(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083720 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/EP02/04005

(22) International Filing Date: 10 April 2002 (10.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

101 18 166.3	11 April 2001 (11.04.2001)	DE
101 30 236.3	22 June 2001 (22.06.2001)	DE
101 55 084.7	9 November 2001 (09.11.2001)	DE
102 01 458.2	16 January 2002 (16.01.2002)	DE

(71) Applicants and
(72) Inventors: **BACHER, Adelbert** [DE/DE]; Königsbergerstr. 74, 85748 Garching (DE). **ROHDICH, Felix** [DE/DE]; Moosburger Str. 16, 85406 Zolling (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ADAM, Petra** [DE/DE]; Engelhofstr. 6, 85386 Dietersheim (DE). **AMSLINGER, Sabine** [DE/DE]; Prof.-Angermair-Ring 24, 85748 Garching (DE). **EISENREICH, Wolfgang** [DE/DE]; Römerhofweg 51a, 85748 Garching (DE). **HECHT, Stefan** [DE/DE]; Engelhofstr. 6, 85386 Dietersheim (DE).

(74) Agent: **HARTZ, Nikolai, F.**; Wächtershäuser & Hartz, Tal 29, 80331 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: INTERMEDIATES AND ENZYMES OF THE NON-MEVALONATE ISOPRENOID PATHWAY

(57) Abstract: The invention provides a protein in a form that is functional for the enzymatic conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate notably in its (*E*)-form of the non-mevalonate biosynthetic pathway to isoprenoids. The invention also provides a protein in a form that is functional for the enzymatic conversion of 1-hydroxy-2-methyl-2butenyl 4-diphosphate, notably in its (*E*)-form, to isopentenyl diphosphate and/or dimethylallyl diphosphate. Further, screening methods for inhibitors of these proteins are provided. Further, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate is provided and chemical and enzymatic methods of its preparation.

WO 02/083720 A2

Intermediates and enzymes of the non-mevalonate isoprenoid pathway

Field of the invention

The present invention relates to cells, cell cultures or organisms or parts thereof for the efficient formation of a biosynthetic product or intermediate or enzyme of a 1-deoxy-D-xylulose 5-phosphate-dependent biosynthetic pathway. Further, the invention relates to vectors for producing them. Further, the invention relates to their use for the formation or production of intermediates or products or enzymes of said biosynthetic pathway as well as to enzymes and intermediates. Further, the invention relates to the screening for inhibitors or enzymes for said biosynthetic pathway.

Background of the invention

The system of biosynthetic pathways in any organism is highly streamlined, whereby a few central trunk pathways branch into a great number of peripheral pathways. The central trunk pathways involve starting materials which are highly integrated. Therefore, central or trunk pathways are highly regulated. At the same time they are crucial for any attempts to interfere with the metabolism of any organism either by an inhibitor or by metabolic engineering.

The isoprenoid pathways are a prime example for this metabolic organisation. They are very long and highly branched, leading to some 30,000 isoprenoid or terpenoid compounds. They all seem to derive from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). They are produced by two alternative trunk pathways (reviewed in Eisenreich et al., 2001).

By the classical research of Bloch, Cornforth, Lynen and co-workers, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) have become established as key intermediates in the biosynthesis of isoprenoids via mevalonate. However, many bacteria, plastids of all plants, and the protozoon *Plasmodium falciparum* synthesize IPP and DMAPP by an alternative pathway via 1-deoxy-D-xylulose 5-phosphate. The discovery of the pathway was mainly based on the incorporation of isotope-labelled 1-deoxy-D-xylulose into the isoprenoid side chain of menaquinones from *Escherichia coli* (Arigoni and Schwarz, 1999).

This mevalonate-independent pathway has so far only been partially explored (Fig. 1). For a better understanding of these aspects of the invention, the pathway shall be briefly explained. It can be divided into three segments:

In a first pathway segment shown in Fig. 1, pyruvate (**1**) is condensed with glyceraldehyde 3-phosphate (**2**) to 1-deoxy-D-xylulose 5-phosphate (DXP) (**3**). Subsequently, DXP is converted into 2*C*-methyl-D-erythritol 4-phosphate (MEP) (**4**) by a two-step reaction comprising a rearrangement and a reduction. This establishes the 5-carbon isoprenoid skeleton.

In the subsequent segment of the mevalonate-independent pathway (Fig. 1), MEP (**4**) is first condensed with CTP to 4-diphosphocytidyl-2C-methyl-D-erythritol (CDP-ME) (**5**) by 4-diphosphocytidyl-2*C*-methyl-D-erythritol synthase (PCT/EP00/07548). CDP-ME (**5**) is subsequently ATP-dependent phosphorylated by 4-diphosphocytidyl-2*C*-methyl-D-erythritol kinase yielding 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate (CDP-MEP) (**6**). The intermediate is subsequently converted into 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate (cMEPP) (**7**) by 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase (PCT/EP00/07548). These three enzymatic steps form a biosynthetic unit which activates the isoprenoid C_5-skeleton for the third pathway segment (Rohdich *et al.*, 1999; Lüttgen *et al.*, 2000; Herz *et al.*, 2000).

Bioinformatic studies (German Patent Application 10027821.3), as well as studies with mutants of *Synechocystis* sp. (Cunningham *et al.*, 2000) and *Escherichia coli* (Campos *et al.*, 2001; Altincicek et al., 2001) demonstrate the involvement of *lytB* and *gcpE* genes in the isoprenoid pathway. However, the function and the reaction catalyzed by the corresponding gene products are still unknown.

Recently, a kinase (XylB) has been described that catalyzes the conversion of 1-deoxy-D-xylulose into 1-deoxy-D-xylulose 5-phosphate at high rates (Wungsintaweekul et al., 2000).

Genes and enzymes participating in further downstream reactions have been described. However, the gene functions, the intermediates, and the mechanisms leading to the products are still unknown.

For numerous pathogenic eubacteria as well as for the malaria parasite *P. falciparum*, the enzymes involved in the non-mevalonate pathway are essential. The intermediates of the mevalonate-independent pathway cannot be assimilated from the environment by pathogenic eubacteria and *P. falciparum*. The enzymes of the alternative isoprenoid pathway do not occur in mammalia which synthesize their isoprenoids and terpenoids exclusively via the mevalonate pathway. Moreover, the idiosyncratic nature of the reactions in this pathway reduces the risk of cross-inhibitions with other, notably mammalian enzymes.

Therefore, enzymes of the alternative isoprenoid pathway seem to be specially suited as

targets for novel agents against pathogenic microorganisms and herbicides. The elucidation of unknown steps and the identification of these targets, e.g. genes and cognate enzymes of these pathways is obligatory for this purpose.

A further source of interest in the non-mevalonate pathway derives from the fact certain pathogens like *Mycobacteria*, *Plasmodia*, *Escherichia* etc. use this pathway to activate γδ T cells (Fournié and Bonneville, 1996). Therefore, γδ T cells likely act as a first line of defense against infections by such pathogens. Intermediates of the non-mevalonate pathway have been suggested to be responsible for γδ T cell activation (Jomaa *et al.*, 1999). Recently, it was show that *E. coli* strains lost the ability to stimulate γδ T cells when the *dxr* or the *gcpE* gene was knocked out (Altincicek *et al.*, 2001).

Moreover, there is a great biotechnological interest in these pathways, since they lead to valuable vitamins and isoprenoid or terpenoid products.

Previous attempts to approach these goals have been hampered by the low rate of biosynthesis along these pathways in wild-type cells studied so far.

Summary of the invention

It is an object of the invention to provide enzymes and nucleic acids coding for said enzymes as well as intermediates for the conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate to isopentenyl diphosphate and/or dimethylallyl diphosphate.

It has surprisingly been found that the intermediate in the conversion of 2C-methyl-D-erythritol 2,4-cyclodiphosphate to isopentenyl diphosphate and/or dimethylallyl diphosphate is 1-hydroxy-2-methyl-2-butenyl 4-diphosphate. This intermediate is formed by an enzyme encoded by *gcpE* as designated in the *E. coli* genome. It has further been found that this enzyme prefers as reductant NADH or NADPH. Further, it has been found that it is promoted by Co^{2+}.

The above intermediate is converted to isopentenyl diphosphate and/or dimethylallyl diphosphate by an enzyme encoded by lytB as designated in the *E. coli* genome. The latter enzyme prefers as reductant NADH or NADPH and FAD as mediator. Further it can be promoted by ions of a metal selected from manganese, iron, cobalt, nickel.

With these findings, the third segment of the trunk non-mevalonate pathway is now

established. The key to these findings is the intermediate 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably in its *E*-form. This establishes the unifying principle of the invention for reactions to and from this intermediate.

Further, it is an object of the invention to provide cells, cell cultures, organisms or parts thereof for the efficient biosynthesis of isoprenoid products or intermediates of the non-mevalonate biosynthetic pathway dependent on 1-deoxy-D-xylulose 5-phosphate production from 1-deoxy-D-xylulose and/or glucose.

The present invention produces a novel *in vivo* system which can be used for the structure elucidation of unknown intermediates and the assignment of biological functions of putative genes or cognate enzymes in the alternative isoprenoid biosynthetic pathway. As an example, the functional assignment of the *gcpE* gene (now designated as *ispG*) and of the *lytB* gene (now designated *ispH*) in the mevalonate-independent pathway of isoprenoid biosynthesis is achieved.

More specifically, said *in vivo* system consists of recombinant *E. coli* strains harbouring vector construct(s) carrying and expressing genes for D-xylulokinase (*xylB*), and genes of further downstream steps of terpenoid biosynthesis, such as *dxs*, *dxr*, and/or *ispD*, and/or *ispE*, and/or *ispF*, and/or *gcpE*, and/or *lytB* from *E. coli*, and/or a carotenoid gene cluster from *Erwinia uredovora*.

In one aspect of the invention, the genetically modified strains can be fed with 1-deoxy-D-xylulose, notably with isotope-labelled 1-deoxy-D-xylulose, which is converted at high rates into the common intermediate of the mevalonate-independent terpenoid pathway, 1-deoxy-D-xylulose 5-phosphate, and into further intermediates of said pathway, like 2*C*-methyl-D-erythritol 4-phosphate, 4-diphosphocytidyl-2*C*-methyl-D-erythritol, 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate, 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, isopentenyl diphosphate, and dimethylallyl diphosphate. Further, feeding with glucose or an intermediate of glycolysis for conversion into said further intermediates of said pathway may be performed.

Said systems are useful for the structure elucidation of hitherto elusive intermediates in the

biosynthetic pathways, for *in vivo* screening of novel antibiotics, antimalarials, and herbicides, and as a platform for the bioconversion of exogenous 1-deoxy-D-xylulose and/or glucose into intermediates and products of the non-mevalonate pathway of terpenoid biosynthesis.
Said systems can also be used for screening chemical libraries for potential herbicides, and/or antimalarials, and/or antimicrobial substances by detecting and measuring the amount of certain intermediates formed *in vivo* in the presence or absence of potential inhibitors of the gene products of mevalonate-independent isoprenoid pathway genes, namely *dxs*, *dxr*, *ispD*, *ispE*, *ispF*, *gcpE,* and *lytB*.

Said system can further be used for the production of higher isoprenoids (e.g. isoprenoids having 10, 15, 20, 30 or 40 carbon atoms) such as carotene, α-tocopherole or vitamins by boosting the bioynthesis of isopentenyl diphosphate and/or dimethylallyl diphosphate via the non-mevalonate pathway, e.g. by using glucose as feeding material. Further feeding materials which may be used are intermediates or products of glycolysis like glyceraldehyde 3-phosphate or pyruvate.

Further, this invention provides novel compounds of formula I (see below), notably 1-hydroxy-2-methyl-2-butenyl 4-diphosphate as well as enzymatic and chemical methods for preparing said compounds. As demonstrated herein, (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate is produced from 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate by the *gcpE* gene product.
It is further demonstrated herein that (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate is converted to dimethylallyl diphosphate and isopentenyl diphosphate by the *lytB* gene product.

Short description of the figures and annexes

Figure 1: Biosynthesis of both isoprenoid precursors, isopentenyl pyrophosphate and dimethylallyl pyrophosphate via the mevalonate-independent pathway.

Figure 2: Scheme of an *Escherichia coli in vivo* system for generating optionally isotopically labelled intermediates of biosynthetic pathways such as the mevalonate-independent isoprenoid biosynthesis, and for the production of higher terpenoids such as carotenoids.

Figure 3: ^1H NMR spectra in D_2O (pH 6) obtained according to Example 25. * indicates impurities.

Figure 4: Preparation of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate according to Example 24. Reagents and conditions were as follows: (a) DHP, PPTS, 25 °C (2.5 h); (b) Ph_3PCHCO_2Et, toluene, reflux (39 h); (c) (1) DIBAH, CH_2Cl_2, –78 °C (3 h), (2) 1 M $NaOH/H_2O$; (d) *p*-TsCl, DMAP, CH_2Cl_2, 25 °C (1 h); (e) $((CH_3CH_2CH_2CH_2)_4N)_3HP_2O_7$, MeCN, 25 °C (2 h); (f), HCl/H_2O pH 1, 25 °C (7 min).

Figure 5: The reaction catalyzed by the *ispH* (formerly *lytB*) gene product.

Figure 6: The reaction catalyzed by the *ispG* (formerly *gcpE*) gene product.

Figure 7: Chemical preparation of 3-formyl-but-2-enyl 1-diphosphate (see example 42).

Annex A: DNA sequence of the vector construct pBSxylBdxr.

Annex B: DNA sequence of the vector construct pBSxylBdxrispD.

Annex C: DNA sequence of the vector construct pBScyclo.

Annex D: DNA sequence of the vector construct pACYCgcpE.

Annex E: DNA sequence of the vector construct pBScaro14.

Annex F: DNA sequence of the vector construct pACYCcaro14.

Annex G: DNA sequence and corresponding amino acid sequence of the *ispG* (formerly *gcpE*) gene from *Escherichia coli*.

Annex H: DNA sequence of the vector construct pBScyclogcpE.

Annex I: DNA sequence of the vector construct pACYClytBgcpE.

Annex J: DNA and corresponding amino acid sequence of the *ispH* (formerly *lytB*) gene from *Escherichia coli*.

Annex K: DNA sequence of the vector construct pBScyclogcpElytB2.

Annex L: DNA and corresponding amino acid sequence of the *ispG* gene (fragment) from *Arabidopsis thaliana*.

Annex M: DNA and corresponding amino acid sequence of the *ispG* (forrmly *gcpE*) gene of *Arabidopsis thaliana*.

Annex N: cDNA sequence of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IspH) from *Arabidopsis thaliana*

Detailed description of the invention

1-Deoxy-D-xylulose 5-phosphate is a common intermediate in the alternative terpenoid pathway via 2*C*-methyl-D-erythritol 4-phosphate. This latter pathway is operative in bacteria, certain protozoa and most significantly also in the plastids of plants, where it is in charge of the biosynthesis of a great many valuable terpenoid products, like natural rubber, carotenoids, menthol, menthone, camphor or paclitaxel. The alternative terpenoid pathway is now intensely studied. But so far only the initial steps from glyceraldehyde 3-phosphate and pyruvate via 1-deoxy-D-xylulose 5-phosphate and 2*C*-methyl-D-erythritol 4-phosphate, 4-diphosphocytidyl-2*C*-methyl-D-erythritol, 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate and 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate (Fig. 1) have been elucidated.

The intermediate 1-deoxy-D-xylulose 5-phosphate is of most crucial significance for a number of commercial purposes:

(1) It may be used as a key intermediate for commercial screening procedures regarding potential inhibitors of downstream enzymes in the biosynthesis of the alternative terpenoid pathway.

(2) It may be used as a key intermediate for the *in vitro* production of terpenoids or of intermediates thereof.

(3) It occurs *in vivo* in the biosynthesis of terpenoids as an enzymatic condensation product of glyceraldehyde 3-phosphate and pyruvate. The latter are central intermediates of the metabolism and obligatory starting materials for numerous biosynthetic pathways. Therefore, it is desirable to generate a high level of 1-deoxy-D-xylulose 5-phosphate *in vivo* from an exogenous source and thus independent from the pools of glyceraldehyde 3-phosphate and pyruvate for boosting the biosynthesis of terpenoids or of intermediates thereof in microorganisms or cell cultures that are either naturally or recombinantly endowed with the pathway of interest without influencing the basic intermediary metabolism of the cells.

(4) 1-Deoxy-D-xylulose 5-phosphate can be generated from 1-deoxy-D-xylulose by the catalytic action of the *xylB* gene product. Using recombinant strains comprising the *xylB* gene the reaction occurs *in vivo* and exogenous 1-deoxy-D-xylulose is converted into intracellular 1-deoxy-D-xylulose 5-phosphate at high rates.

(5) 1-DXP can be generated fro glucose by the catalytic action of glycolytic enzymes and DXP-synthase. Using recombinant strains comprising the *dxs* gene, the reaction occurs *in vivo* and exogeneous glucose is converted to intracellular 1-DXP at high rates.

It is an aspect of the invention to use 1-deoxy-D-xylulose as a precursor in order to boost the rates of biosynthesis of 1-deoxy-D-xylulose 5-phosphate-dependent pathways. 1-Deoxy-D-xylulose can be prepared by various published procedures (Blagg and Poulter,1999; Kennedy et al., 1995; Piel and Boland, 1997; Shono et al., 1983; Giner, 1998).

It is an aspect of the present invention to use 1-deoxy-D-xylulose in various isotopically labelled forms. It may be labelled by radioactive isotopes or non-radioactive isotopes of C (^{13}C or ^{14}C), H (D or T) or 0 ($^{17}0$ or $^{18}0$) in any combination.

Isotope-labelled 1-deoxy-D-xylulose may be prepared enzymatically using 1-deoxy-D-xylulose 5-phosphate synthase of *Bacillus subtilis* and commercially available glycolytic enzymes and phosphatase from isotope-labelled glucose and/or pyruvate (PCT/EP00/07548).
1-Deoxy-D-xylulose may be used as a free acid or as a salt, preferably as an alkaline (e. g., lithium, sodium, potassium) salt or as an ammonium or amine salt.

It is an aspect of the present invention to use recombinant cells, cell cultures, or organisms or parts thereof for the formation of biosynthetic products or intermediates or enzymes or for the screening for antimicrobials, antimalarials or herbicides.

For carrying out the present invention various techniques in molecular biology, microbiology and recombinant DNA technology are used which are comprehensively described in Sambrock et al., Molecular Cloning, second edition, Cold Spring Harbor Laboratory Press, Cold Sprind Harbor, New York; in DNA Cloning: A Practical Approach, Vol. 1 and 2, 1985 (D. N. Glover, ed.); in Oligonucleotide Synthesis, 1984 (M. L. Gait, ed.); and in Transcription and Translation (Hames and Higgins, eds.).

Nucleic acids

The present invention comprises nucleic acids which include prokaryotic, protozoal and plant sequences and derived sequences. A derived sequence relates to a nucleic acid sequence corresponding to a region of the sequence or orthologs thereof or complementary to "sequence-conservative" or "function-conservative" variants thereof.
Sequences may be isolated by well known techniques or are commercially available (Clontech, Palto Alto, CA; Stratagene, LaJolla, CA). Alternatively, PCR-based methods can be used for

amplifying related sequence from cDNA or genomic DNA.

The nucleic acids of the present invention comprise purine and pyrimidine containing polymers in various amounts, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribonucleotides. The nucleic acids may be isolated directly from cells. Alternatively, PCR may be used for the preparation of the nucleic acids by use of chemical synthesized strands or by genomic material as template. The primers used in PCR may be synthesized by using the sequence information provided by the present invention or from the database and additionally may be constructed with optionally new restriction sites in order to ease the cloning in a vector for recombinant expression.

The nucleic acids or the present invention may be flanked by natural regulation sequences or may be associated with heterologous sequences, including promoter, enhancer, response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3' noncoding regions or similar. The nucleic acids may be modified on basis of well known methods. Non-limiting examples for these modifications are methylations, "Caps", substitution of one or more natural nucleotides with an analogue, and internucleotide modification, i.e. those with uncharged bond (i.e. methylphosphonates, phosphotriester, phosphoramidates, carbamates, etc.) and with charged bond (i.e. phosphorothiactes, etc.). Nucleic acids may carry additional kovalent bound units such as proteins (i.e. nucleases, toxins, antibodies, signalpeptides, poly-L-lysine, etc.), intercalators (i.e. acridine, psoralene, etc.), chelators (i.e. metals, radioactive metals, iron, oxidative metals, etc.) and alkylators. The nucleic acids may be derived by formation of a methyl- or ethylphosphotriester bond or of a alkylphosphoramidate bond. Further, the nucleic acids of the present invention may be modified my labeling, which give an either directly or indirectly detectable signal. Examples for these labeling include radioisotopes, fluorescent molecules, biotin and so on.

Vectors

The invention provides nucleic acid vectors, which comprise the sequences provided by the present invention or derivatives thereof. Various vectors, including plasmids or vectors for fungi have been described for the replication and/or expression in various eucaryotic and procaryotic hosts. High copy replication vectors are preferred for the purposes of the invention. Non-limiting examples include pKK plasmids (Clontech), pUC plasmids (Invitrogen, San Diego, CA), pET plasmids (Novagen, Inc., Madison, WI) or pRSET or pREP (Invitrogen) and various suitable host cells on basis of well known techniques. Recombinant cloning vectors comprise

often more than one replication system for the cloning and expression, one or more marker for the selection in the host; i.e. antibiotic resistance and one or more expression cartridge. Suitable hosts may be transformed/transfected/infected by a method as suitable including electroporation, $CaCl_2$-mediated DNA incorporation, tungae infection, microinjection, microbombardment or other established methods.

Suitable hosts include bacteria, archaebacteriae, fungi, notable yeast, plants, notably *Arabidopsis thaliana, Mentha piperita or Taxus sp.* and animal cells, notably mammalian cells. Most important are *E. coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis, Schizosaccharomyces pombe*, SF9 cells, C129 cells, 293 cells, *Neurospora*, and CHO cells, COS cells, HeLa cells and immortalized myeloid and lymphoid mammalian cells. Preferred replication systems include M13, ColE1, SV40, baculovirus, lambda, adenovirus and so on. A great number of transcription, initiation (including ribosomal binding sites) and termination regulation regions have been isolated and there efficiency for the transcription and translation of heterologous proteins has been demonstrated in various hosts. Examples for these regions, methods for the isolation, the way for using are well known. Under suitable conditions for expression host cells may be used as source for the recombinant synthesized proteins.

Expression systems

Preferable vectors may include a transcription element (that is a promoter), functionally connected with the enzyme domain. Optionally, the promoter may include parts of operator region and/or ribosomal binding sites. Non-limiting examples for bacterial promoters, which are compatible with *E. coli*, include: *trc* promoter, β-lactamase (penicillinase) promoter; lactose promoter, tryptophan (*trp*) promoter, arabinose BAD operon-promoter, lambda-derived P1 promoter and N gene ribosomal binding site and the hybrid Tac promoter, derived from sequences of *trp* and *lac* UV5 promoters. Non-limiting examples for yeast promoters include 3-phosphoglycerate kinase promoter, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter, galactokinase (GALI) promoter, galactoepimerase promoter and alcoholdehydrogenase (ADH) promoter. Suitable promoters for mammalian cells include without limiting viral promoters such as i.e. simian virus 40 (SV40), rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also need terminator sequences and poly-A sequences and enhancer sequences, which may increase the expression. Sequences, which amplify the genes, may also be preferred. Further on,

sequences may be included, which ease the secretion of the recombinant protein from the cell, which may be but non-limiting a bacterial, yeast or animal cell, such as i.e. a secretion signal sequence and/or prehormon sequence.

It is an important aspect of the invention that the combined recombinant endowment with *xylB* and other gene(s) of the alternative C5-isoprenoid pathway and optionally gene(s) for higher isoprenoids or terpenoids boost(s) these pathways. Preferably, *xylB* is combined with complete sets of genes to convert 1-deoxy-D-xylulose 5-phosphate into the desired intermediate or end products. For intermediates in the C5-isoprenoid pathway, cells are preferably endowed with one of the combinations of genes given in claim 76.

For the genes cited herein, the common *E. coli* designation were used. Other genes from *E. coli* or from other organisms (orthologous genes) may also be used if they have the same functions (function-conservative genes), notably if their gene products catalyze the same reaction. Further, deletion or insertion variants or fusions of these genes with other genes or nucleic acids may be used, as long as these variants are function-conservative. The above genes may be derived from bacteria, protozoa, or from higher or lower plants,

It is another important aspect of the invention that the function of *gcpE* as following immediately downstream from *ispF* has been determined. Our findings show that the *gcpE* gene product is involved in the formation of the novel compound 1-hydroxy-2-methyl-2-butenyl 4-diphosphate from 2C-methyl-D-erythritol 2,4-cyclodiphosphate. Therefore, we rename *gcpE* in *ispG*.

In a further aspect of the invention it was shown that the gene product of *gcpE* is involved in the formation of the *E*-isomer of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate from 2C-methyl-D-erythritol 2,4-cyclodiphosphate by comparison with chemically synthesized (*E*)- and (*Z*)-isomers of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate. Therefore, this invention further pertains to the (*E*) and (*Z*) isomers of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate salts or protonated forms thereof.

It is another important aspect of the invention that the function of *lytB* has been determined as following immediately downstream from *ispG*. Therefore, it is renamed *ispH*. It is our finding

that ispH is involved in the conversion of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate into isopentenyl 4-diphosphate and/or dimethylallyl 4-diphosphate.

It should be understood that „1-hydroxy-2-methyl-2-butenyl 4-phosphate" and „1-hydroxy-2-methyl-2-butenyl 4-diphosphate" comprise the free phosphoric and diphosphoric acids, respectively, and the singly or multiply deprotonated forms thereof, i.e. salts which may be salts of any cation (including Na, K, NH_4^+, Li, Mg, Ca, Zn, Mn, and Co cations). The protonation state of (di)phosphates and phosphate derivatives or their conjugated acids in aqueous solution depends on the pH value of the solution, as is known to persons skilled in the art. The same applies to other phoshates or phosphate derivatives.

In another aspect of the invention, (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate has been successfully incorporated into the lipid soluble fraction of *Capsicum annuum* chromoplasts. A ^{14}C label of this compound was incorporated into the geranylgeraniol, β-carotene, phytoene and phytofluene fractions of *C. annuum* chromoplasts establishing (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate as intermediate of the non-mevalonate pathway downstream from 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and upstream from isopentenyl diphosphate.

It is another aspect of the invention that *xylB* can be combined with *gcpE* and optionally other genes of the alternative C5 isoprenoid pathway and/or of the higher isoprenoid pathways in vector(s) for recombinant engineering.

As a consequence of our findings regarding *gcpE* (now *ispG*) it follows that the gene *lytB* operates downstream of *gcpE* and thus in service of the conversion of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate to IPP and/or DMAPP. Therefore, it is another aspect of the invention to combine the gene *lytB* with *xylB* and optionally other genes of the common C5-isoprenoid pathway or of a higher isoprenoid pathway.

Our finding allows the efficient formation or production of intermediates or products of the isoprenoid pathway with any desired labelling, notably the following intermediates:
2C-methyl-D-erythritol 4-phosphate; 4-diphosphocytidyl-2C-methyl-D-erythritol; 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate; 2C-methyl-D-erythritol 2,4-

cyclodiphosphate; 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, isopentenyl diphosphate; dimethylallyl diphosphate.

The formation of end products of the terpenoid pathway (e. g., β-carotene, zeaxanthine, paclitaxel, menthol, menthone, cannabinoids), may be boosted following the process of the invention.

The strains harbouring the recombinant plasmids can be cultivated in conventional culture media, preferably in terrific broth medium, at 15 to 40 °C. The preferred temperature is 37 °C. The *E. coli* strains are induced with 0.5 to 2 mM isopropyl-β-D-thiogalactoside (IPTG) at an optical density at 600 nm from 0.5 to 5. The cells are incubated after addition of 1-deoxy-D-xylulose at a concentration of 0.001 mM to 1 M preferably at a concentration of 0.01 to 30 mM for 30 min to 15 h, preferably 1 to 5 h.

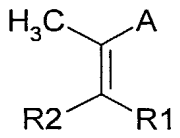
It has been found that the process of producing isoprenoid intermediates or products by the genetically engineered organisms of the invention can be boosted by supplying a source for CTP, for example cytidine and/or uridine and/or cytosine and/or uracil and/or ribose and/or ribose 5-phosphate and/or any biosynthetic precursors of CTP at a concentration of 0.01 to 10 mM, preferably at a concentration of 0.3 to 1 mM, and/or by supplying a source for phosphorylation activity, for example glycerol 3-phosphate and/or phosphoenolpyruvate and/or ribose 5-phosphate at a concentration of 0.1 to 100 mM, preferably at a concentration of 0.5 to 10 mM and/or inorganic phosphate and/or inorganic pyrophosphate at a concentration of 1 to 500 mM, preferably at a concentration of 10 to 100 mM and/or any organic phosphate and/or pyrophosphate, and/or by supplying a source for reduction equivalents, for example 0.1 to 1000 mM, preferably 10 to 1000 mM, lactate and/or succinate and/or glycerol and/or glucose and/or lipids at a concentration of 0.1 to 100 mM, preferably at a concentration of 0.5 to 10 mM. A particularly efficient production process is specified in claims 72 and 80 to 84.

This process can also be used with great advantages for screening for inhibitors of the enzymes involved or of downstream enzymes, dependent on the choice of the isoprenoid intermediate or product for detection. The enzymes *dxs, dxr*, *ispD*, *ispE, ispF, ispG* (formerly *gcpE*) and *ispH* (formerly *lytB)* do not occur in animals. Therefore inhibitors against *dxs*, *dxr*, *ispD*, *ispE, ispF, ispG* (formerly *gcpE*) and *ispH* (formerly *lytB)* have great value as (a) herbicides against weed plants or algae; (b) antibiotic agents against pathogenic bacteria; (c) agents against protozoa, like *Plasmodium falciparum*, the causative pathogen of malaria.

The activity of the said enzymes can be detected (in the presence or absence of a potential inhibitor) by measuring either the formation of a product or the consumption of an intermediate, preferably by TLC, HPLC or NMR.

With the finding that 1-hydroxy-2-methyl-2-butenyl 4-diphosphate is an intermediate of the non-mevalonate terpenoid pathway we have aquired essential determinants of the structure of inhibitors. Namely, the structures of a subset of inhibitors should be similar to at least a portion of the starting compound or the product or the transition state between the starting compound e.g. 2C-methyl-D-erythritol 2,4-cyclodiphosphate and the product e. g. 1-hydroxy-2-methyl-2-butenyl 4-diphosphate.

This invention discloses novel compounds, or salts thereof, of the following formula I:



(I)

whereby R^1 and R^2 are different from each other and one of R^1 and R^2 is hydrogen and the other is selected from the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$, $-CH_2-O-PO(OH)_2$, and $-CH_2OH$, and whereby A stands for $-CH_2OH$ or -CHO. These compounds may be isotope-labelled.

In formula I, A preferably stands for $-CH_2OH$.

Among R^1 and R^2, R^1 is preferably hydrogen and R^2 is preferably selected from the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$ and $-CH_2-O-PO(OH)_2$.

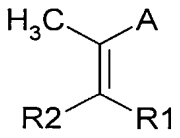
In the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$ and $-CH_2-O-PO(OH)_2$, $-CH_2-O-PO(OH)-O-PO(OH)_2$ is preferred.

If a compound of formula I is a salt, it may e.g. be a lithium, sodium, potassium, magnesium, ammonium, manganese salt. These salts may derive from a single or from multiple deprotonations from the (di)phosphoric acid moiety.

The novel compounds disclosed herein are useful for various applications e.g. for screening for genes, enzymes or inhibitors of the biosynthesis of isoprenoids or terpenoids, either *in vitro* in the presence of an electron donor or *in vivo*.

This invention further provides a process for the chemical preparation of a compound of

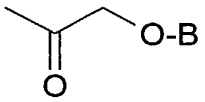
formula I or a salt thereof:



(I)

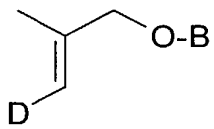
wherein A represents $-CH_2OH$ and R^1 and R^2 are different from each other and one of R^1 and R^2 is hydrogen and the other is $-CH_2-O-PO(OH)-O-PO(OH)_2$, $-CH_2-O-PO(OH)_2$ or $-CH_2-OH$ by the following steps:

(a) converting a compound of the following formula (II):

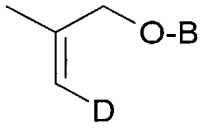


(II)

wherein B is a protective group into a compound of the following formula (III) or (IV):



(III)



(IV)

by a Wittig or Horner reagent, wherein the group D is a precursor group convertible reductively to a $-CH_2-OH$ group;

(b) reductively converting group D to a $-CH_2-OH$ group;

(c) optionally converting group $-CH_2-OH$ obtained in step (b) into $-CH_2-O-PO(OH)-O-PO(OH)_2$ or $-CH_2-O-PO(OH)_2$ or salts thereof in a manner knwon per se;

(d) optionally conversion to a desired salt;

(e) removing the protective group B.

In the above process, said protective group B may be any group that allows to regenerate an hydroxy group at the position it is attached to. Said protective group B is preferably stable under the conditions of step (a) to step (d). Said protective group B is removed in step (e) of said process in order to generate a hydroxy group. Protective groups for hydroxy groups are known to the skilled person. Group B may for example form an acetal group together with the

remaining moiety of the compound of formula (II), (III) or (IV). Acetals can be hydrolysed under acidic conditions. Most preferably, group B is a 2-tetrahydropyranyl group.

In the above process, said group D is a precursor group convertible reductively to a $-CH_2-OH$ group. Group D may be a derivative of a carbon acid. Examples of such a group include alkoxycarbonyl and aminocarbonyl groups. Said aminocarbonyl groups may be substitued at the amino group with one or two alkyl groups. It is most preferred to use alkoxycarbonyl groups. The alkyl group of said alkoxycarbonyl groups or said alkyl groups of said aminocarbonyl groups may be a linear or branched alkyl groups which may be singly or multiply substituted. Preferred are C_1-C_6 alkyl groups like methyl, ethyl, propyl, butyl, pentyl or hexyl groups. Most preferred are methyl or ethyl groups. The most preferred example of said group D is an ethoxycarbonyl group.

Said compound of formula (II) may be prepared by protecting the hydroxy group of hydroxy acetone with said group B. If group B is a tetrahydropyranyl group, the compound of formula (II) may be prepared from hydroxy acetone and 3,4-dihydro-2H-pyran, preferably employing pyridinium toluene-4-sulfonate as a catalyst. A specific method for preparing acetonyl tetrahydropyranyl ether is described in example 24.

In step (a) of said process, the compound of formula (II) is converted to a compound of formula (III) or (IV) by a Wittig or a Horner reagent. Wittig-type reactions and reagents are known to skilled persons (see e.g. Watanabe *et al.* 1996 and references cited therein). Common Wittig reagents to be used for the above process are methylen-triphenylphosphoranes which may be substituted at the methylene group. For the above process of this invention, a methylen-triphenylphosphorane is employed which is substituted with the above-defined group D at the methylene group. Such Wittig reagents are commercially available or can be prepared according to known methods.
The olefin produced in step (a) may be formed as a mixture of the cis/trans isomers of formulas (III) and (IV). If one of said isomers is preferred, it may be enriched or separated from the other isomer by methods known in the art, preferably by chromatography. Alternatively, a separation of said isomers may be carried out after one of the following steps (b) to (e).

In step (b) of the above process, group D of the compound of formula (III) or (IV) or a mixture of said compounds is reductively converted to a $-CH_2-OH$ group. Various methods are known

in the art to perform such a reduction. Conditions are chosen such that group D is reduced whereas the olefin moiety is not. Examples for reductants to be used in this step are molecular hydrogen or metal hydrides. Examples for useful metal hydrides include boron hydrides like sodium borohydride, aluminium hydrides like lithium aluminium hydride or diisobutyl aluminiumhydride (DIBAH), alkali metal or metal earth hydrides like sodium hydride or calcium hydride. Aluminium hydrides are preferred. A specific example for carrying out step (b) is described in example 24.

If the desired end product of said process is a compound of formula (I), wherein R^1 or R^2 is $-CH_2-OH$, the compound or mixture of compounds obtained in step (b) may be directly subjected to step (d) or step (e). Preferably, it is subjected to step (e) for removing protective group B. If the desired end product of said process is a compound of formula (I), wherein R^1 or R^2 is $-CH_2-O-PO(OH)-O-PO(OH)_2$ or $-CH_2-O-PO(OH)_2$, compound or mixture of compounds obtained in step (b) is subjected to step (c) of said process for converting $-CH_2-OH$ group obtained in step (b) into a $-CH_2-O-PO(OH)-O-PO(OH)_2$ or a $-CH_2-O-PO(OH)_2$ group.

Step (c) may be carried in several ways which are known to the skilled person. Step (c) may comprise substituting the hydroxy group of said $-CH_2-OH$ group obtained in step (b) by a leaving group. Step (c) may comprise converting said $-CH_2-OH$ group to a $-CH_2$-halide group by a halogenating agent. A sulfuric, sulfonic or phosphoric acid halogenide may be employed as halogenating agent. Tosyl chloride is most preferred. Said halide may be fluoride, chloride, bromide or iodide, preferably chloride. The compound carrying said $-CH_2$-halide group is preferably isolated. Said leaving group may further be created by reacting said $-CH_2-OH$ group obtained in step (b) with a sulfonic acid halide, preferably tosyl chloride.
Said intermediate having said leaving group may then be reacted with phosphoric or diphosphoric acid or singly or multiply deprotonated forms thereof. Preferably an alkylammonium salt of phosphoric or diphosphoric acid is used, more preferably a tetraalkylammonium salt, and most preferably a tetra-butylammonium salt. A polar aprotic solvent is preferred for this reaction. Preferably, the compound or mixture of compounds obtained is purified according to standard procedures. A specific example for carrying out step (c) is described in example 24.

In step (d), the compound or mixture of compounds obtained in step (c) may be converted to a desired salt. Methods for carrying out step (d) are well known. Such methods may comprise

adjusting the pH of an aqueous solution with an appropriate acid or salt to a desired pH value.

In step (e), the protective group B of a compound obtained in one of steps (b) to (d) is removed in order to obtain a compound of formula (I) wherein A is $-CH_2-OH$. The method for removing a protective group depends on the type of the protective group. Such methods are well known. If the protective groups forms an acetal, removing said protecting group may be achieved by acid hydrolysis (see example 24).

This invention provides protein in a form that is functional for the enzymatic conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate notably in its (*E*)-form, preferably in the presence of NADH and/or NADPH and/or in the presence of Co^{2+}. Said enzyme preferably has a sequence encoded by the *ispG* (formerly *gcpE*) gene of *E. coli* or a function-conservative homologue of said sequence, i.e. said homologue is capable of performing the same function as said protein. For many applications of said protein, it may be expressed and purified as a fusion protein, notably a fusion with maltose binding protein. In this way, enzymatically active protein may be readily obtained.

This invention further provides a protein in a form that is functional for the enzymatic conversion of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably in its (*E*)-form, to isopentenyl diphosphate and/or dimethylallyl diphosphate. Said protein preferably requires FAD and NAD(P)H for said functionality. Further, said protein may require a metal ion selected from the group of manganese, iron, cobalt, or nickel ion. Said protein preferably has a sequence encoded by the *ispH* (formerly *lytB*) gene of *E. coli* or a function-conservative homologue of said sequence. For many applications of said protein, it may be expressed and purified as a fusion protein, notably a fusion with maltose binding protein. In this way, enzymatically active protein may be readily obtained.

The above proteins may be plant proteins, notably from *Arabidopsis thaliana*, bacterial proteins, notably from *E. coli*, or protozoal proteins, notably from *Plasmodium falciparum*.

The invention further provides a purified isolated nucleic acid encoding one or both of the above proteins with or without introns. Further, the invention provides a DNA expression vector containing the sequence of said purified isolated nucleic acid.

The invention further provides cells, cell cultures, organisms or parts thereof recombinantly endowed with the sequence of said purified isolated nucleic acid or with said DNA expression vector, wherein said cell is selected from the group consisting of bacterial, protozoal, fungal, plant, insect and mammalian cells. Said cells, cell cultures, organisms or parts thereof may further be endowed with at least one gene selected from the following group: *dxs*, *dxr*, *ispD* (formerly *ygbP*); *ispE* (formerly *ychB*); *ispF* (formerly *ygbB*) of *E. coli* or a function-conservative homologue thereof, or a function-conservative fusion, deletion or insertion variant of any of the above genes.

The invention further provides cells, cell cultures, or organisms or parts thereof transformed or transfected for an increased rate of formation of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably in its (*E*)-form, compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection. The transformation or transfection preferably comprises endowment with the *gcpE* gene of *E. coli* or with a function-conservative homologue from an other organism, e.g. plant or protozoal organism.

The invention also provides cells, cell cultures, or organisms or parts thereof transformed or transfected for an increased rate of conversion of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate to isopentenyl diphosphate and/or dimethylallyl diphosphate compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection. The transformation or transfection preferably comprises endowment with the *lytB* gene of *E. coli* or with a function-conservative homologue from an other organism, e.g. plant or protozoal organism.

The invention provides also cells, cell cultures, or organisms or parts thereof transformed or transfected for an increased expression level of the protein of one of claims 1 to 4 and/or the protein of one of claims 5 to 8 compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection.

Moreover, the invention provides a method of altering the expression level of the gene product(s) of *ispG* and/or *ispH* or function-conservative homolgues from other organisms or variants thereof in cells comprising

(a) transforming host cells with the *ispG* and/or *ispH* gene,

(b) growing the transformed host cells of step (a) under conditions that are suitable for the

efficient expression of *ispG* and/or *ispH*, resulting in production of altered levels of the ispG and/or ispH gene product(s) in the transformed cells relative to expression levels of untransformed cells.

Furthermore, the invention provides a method of identifying an inhibitior of an enzyme functional for the conversion of 2C-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably its *E*-form, of the non-mevalonate isoprenoid pathway by the following steps:

(a) incubating a mixture containing said enzyme with its, optionally isotope-labeled, substrate 2C-methyl-D-erythritol-2,4-cyclodiphosphate under conditions suitable for said conversion in the presence and in the absence of a potential inhibitor,

(b) subsequently determining the concentration of 2C-methyl-D-erythritol 2,4-cyclodiphosphate and/or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, and

(c) comparing the concentration in the presence and in the absence of said potential inhibitor.

Furthermore, the invention provides a method of identifying an inhibitior of an enzyme functional for the conversion of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably its *E*-form, to isopentenyl diphosphate or dimethylallyl diphosphate of the non-mevalonate isoprenoid pathway by the following steps:

(a) incubating a mixture containing said enzyme with its, optionally isotope-labeled, substrate 1-hydroxy-2-methyl-2-butenyl 4-diphosphate under conditions suitable for said conversion in the presence and in the absence of a potential inhibitor, whereby said mixture preferably contains FAD,

(b) determining the concentration of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate or dimethylallyl diphosphate, and

(c) comparing the concentration in the presence and in the absence of said potential inhibitor.

The above methods of identifying an inhibitior are preferably carried out by following the consumption of NADPH or NADH making use of its characteristic absorbance spectrum. Alternatively, the fluorescence of NADH or NADPH can be followed when excited around 340 nm. The above methods of identifying an inhibitior may advantageously be performed as high-

throughput screening assays for inhibitors, notably in combination with photometric detection of the consumption of NADH or NADPH. Further, one or more flavin analogues (e.g. FAD, FMN) may be added to the incubation mixtures in said methods, preferably in catalytic amounts. Most preferred is the addition of FAD. Said enzymes may be employed in said methods as fusion proteins with maltose binding protein (examples 38 to 41, 44, 45), which allows straightforward expression and purification of said enzymes in enzymatically active form. Further embodiments of said methods of identifying are defined in the subclaims to these methods.

It is known that intermediates of the non-mevalonate pathway are responsible for γδ T cell activation by various pathogenic bacteria. γδ T cell activation is followed by T cell proliferation, secretion of cytokines and chemokines and is very likely crucial for regulating the immune response following pathogen infection (Altincicek *et al.*, 2001 and references cited therein). Recently, it was shown that *E. coli* strains lost the ability to stimulate γδ T cells when the *dxr* or the *gcpE* gene was knocked out, strongly indicating that an intermediate downstream of *gcpE* and upstream of isopentenyl pyrophosphate exhibits the most potent antigenic activity (Altincicek *et al.*, 2001). However, the intermediate produced by the *gcpE* gene product in the pathway has been unknown. Herein, this intermediate has surprisingly been identified as an hitherto unprecedented compound, which opens up a whole range of novel applications for this compound.

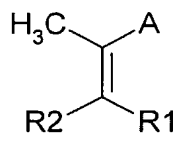
The compounds of formula I can be used as immunomodulatory or immunostimulating agents, e.g. for activating γδ T cells. Immunomodulation via γδ T cell activation by said compounds may prove useful not only to support combat against pathogens but for various conditions for which a stimulation of the immune system is desirable. The novel compounds of the invention may therefore be used for medical treatment of pathogen infections. Such a treatment stimulates the activity of the immune system against the pathogen. Preferably, the compound wherein R^1=H and/or A is $-CH_2OH$ is used for this application. Alternatively, the oxidation product with A = CHO may prove to be highly active. Among the compounds of formula I, the one with the highest or most suitable γδ T cell stimulating activity may be selected in a test system known in the art (e.g. that described by Altincicek *et al.*, 2001). Importantly, since the compounds of the invention do not act as antibiotics, development of resistancies is not a problem for the method of treatment disclosed herein.

In an advantageous embodiment, said compounds may be combined with an antibiotically active compound for treating a pathogen infection. Such a treatment combines the advantages of inhibiting pathogen proliferation by an antibiotic and stimulating the immune system against the pathogen resulting in a much faster and more efficient treatment. Such an antibiotically active compound may be a bacteriostatic antibiotic (e.g. tetracyclines).

Therefore, the novel compounds of this invention may be used for the preparation of a medicament. The invention further pertains to a pharmaceutical composition containing a compound of formula I and a pharmaceutically acceptable carrier. Said pharmaceutical composition may further contain an antibiotically active compound as mentioned above.

This invention further comprises antibodies against the compounds of formula I. Said antibodies may be polyclonal or monoclonal and may be raised according to conventional techniques. Raising such antibodies will comprise coupling of a compound of formula I has hapten to a macromolecular carrier like a protein in order to be immunogenic. Such an immunogenic compound of formula I may further be used as a vaccine.
The antibodies of the invention may be used for detecting a compound of formula I. Since said compounds are produced by organisms having the non-mevalonate isoprenoid pathway, such organisms may be detected using said antibodies. Preferably, such organisms may be detected in body fluids in a diagnostic method, thereby indicating an infection by a pathogen having the non-mevalonate pathway. A positve result in such a diagnostic method may at the same time indicate possible treatment by the compounds of the invention.
When an antibody of the invention is used for detecting a compound of formula I, it is preferably labelled to allow photometric detection and/or immobilized to a support. Such methods are well-known in the art.

This invention further provides a process for the chemical preparation of a compound of formula I or a salt thereof (see Fig. 7):



(I)

wherein A represents $-CH_2OH$ or -CHO, R^1 is hydrogen, and R^2 is $-CH_2$-O-PO(OH)-O-

$PO(OH)_2$, $-CH_2-O-PO(OH)_2$ or $-CH_2-OH$ by the following steps:

(a) converting 2-methyl-2-vinyl-oxiran into 4-chloro-2-methyl-2-buten-1-al;

(b) converting 4-chloro-2-methyl-2-buten-1-al to its acetal;

(c) substituting the chlorine atom in the product of step (b) by a hydroxyl group, a phosphate group or a pyrophosphate group;

(d) hydrolysing the acetal obtained in step (c) to produce an aldehyde group;

(e) optionally converting the aldehyde group of the product of step (d) to a $-CH_2OH$ group.

Preferred embodiments of this process are defined in the subclaims and are exemplified in example 42.

The invention will now be described in detail with reference to specific examples.

Example 1

Construction of a vector carrying the *xylB* gene of *Escherichia coli* capable for transcription and expression of D-xylulokinase

Chromosomal DNA from *Escherichia coli* strain XL1-Blue (Bullock et al. 1987; commercial source: Stratagene, LaJolla, CA, USA) is isolated according to a method described by Meade et al. 1982.

The *E. coli* ORF *xylB* (accession no. gb AE000433) from base pair (bp) position 8596 to 10144 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CCGTCGGAATTCGAGGAGAAATTAACCATGTATATCGGGATAGATCTTGG-3', 10 pmol of the primer 5'-GCAGTGAAGCTTTTACGCCATTAATGGCAGAAGTTGC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec, Seraing, Belgium) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 75 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden, Germany).

1.0 µg of the vector pBluescript SKII$^-$ (Stratagene) and 0.5 µg of the purified PCR product are digested with *Eco*RI and *Hind*III in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (New England Biolabs, Frankfurt am Main, Germany (NEB)) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 20 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBSxylB. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells according to a method described by Dower *et al.*, 1988. The plasmid pBSxylB is isolated with the plasmid

isolation kit from Qiagen.

The DNA insert of the plasmid pBSxylB is sequenced by the automated dideoxynucleotide method (Sanger *et al.*, 1992) using an ABI Prism 377™ DNA sequencer from Perkin Elmer (Norwalk, USA) with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions (Foster city, USA). It is identical with the DNA sequence of the database entry (gb AE000433).

Example 2

Construction of a vector carrying the *xylB* and *dxr* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase and DXP reductoisomerase

The *E. coli* ORF *dxr* (accession no. gb AE000126) from base pair (bp) position 9887 to 11083 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CTAGCCAAGCTTGAGGAGAAATTAACCATGAAGCAACTCACCATTCTGG-3', 10 pmol of the primer 5'-GGAGATGTCGACTCAGCTTGCGAGACGC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 75 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.2 µg of the vector pBSxylB (Example 1) and 0.6 µg of the purified PCR product are digested with *Hind*III and *Sal*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 18 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBSxylBdxr. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBSxylBdxr is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBSxylBdxr is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000126).
The DNA sequence of the vector construct pBSxylBdxr is shown in Annex A.

Example 3

Construction of a vector carrying the *xylB*, *dxr* and *ispD* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase and CDP-ME synthase

The *E. coli* ORF *ispD* (accession no. gb AE000358) from base pair (bp) position 6754 to 7464 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CCGGGAGTCGACGAGGAGAAATTAACCATGGCAACCACTCATTTGGATG-3', 10 pmol of the primer 5'-GTCCAACTCGAGTTATGTATTCTCCTTGATGG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 30 sec at 94 °C, 30 sec at 50 °C and 45 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.5 µg of the vector pBSxylBdxr (Example 2) and 0.8 µg of the purified PCR product are digested with *Sal*I and *Xho*I in order to produce DNA fragments with overlapping ends. The

restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 12 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBSxylBdxrispD. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBSxylBdxrispD is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBSxylBdxrispD is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000126).
The DNA sequence of the vector construct pBSxylBdxrispD is shown in Annex B.

Example 4

Construction of a vector carrying the *xylB*, *dxr*, *ispD* and *ispF* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase, CDP-ME synthase, and cMEPP synthase

The *E. coli* ORF's *ispD* and *ispF* (accession no. gb AE000358) from base pair (bp) position 6275 to 7464 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CCGGGAGTCGACGAGGAGAAATTAACCATGGCAACCACTCATTTGGATG-3', 10 pmol of the primer 5'-TATCAACTCGAGTCATTTTGTTGCCTTAATGAG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 75 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.4 µg of the vector pBSxylBdxr (Example 2) and 0.7 µg of the purified PCR product are digested with *Sal*I and *Xho*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 18 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBSxylBdxrispDF. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBSxylBdxrispDF is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBSxylBdxrispDF is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000126).

Example 5

Construction of a vector carrying the *xylB*, *dxr*, *ispD*, *ispE* and *ispF* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase, CDP-ME synthase, CDP-ME kinase and cMEPP synthase

The *E. coli* ORF *ispE* (accession no. gb AE000219) from base pair (bp) position 5720 to 6571 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GCGAACCTCGAGGAGGAGAAATTAACCATGCGGACACAGTGGCCC-3', 10 pmol of the primer 5'-CCTGACGGTACCTTAAAGCATGGCTCTGTGC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.2 µg of the vector pBSxylBdxrispDF (Example 4) and 0.6 µg of the purified PCR product are digested with *Xho*I and *Kpn*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 15 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScyclo. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScyclo is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScyclo is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000219). The DNA sequence of the vector construct pBScyclo is shown in Annex C.

Example 6

Construction of a vector carrying the *gcpE* gene of *Escherichia coli* capable for its transcription and expression

The *E. coli* ORF *gcpE* (accession no. gb AE000338) from base pair (bp) position 372 to 1204 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CGTACCGGATCCGAGGAGAAATTAACCATGCATAACCAGGCTCCAATTC-3', 10 pmol of the primer 5'-CCCATCGTCGACTTATTTTTCAACCTGCTGAACGTC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of

100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.0 µg of the vector pACYC184 (Chang and Cohen 1978, NEB) and 0.7 µg of the purified PCR product are digested with *Bam*HI and *Sal*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 20 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYCgcpE. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYCgcpE is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYCgcpE is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000338).
The DNA sequence of the vector construct pACYCgcpE is shown in Annex D.

Example 7

Construction of vectors carrying a carotenoid operon from *Erwinia uredovora* capable for the *in vivo* production of β-carotene

The open reading frames *crtY*, *crtI* and *crtB* of a carotenoid operon from *Erwinia uredovora* (accession no. gb D90087) from base pair (bp) position 2372 to 6005 is amplified by PCR

using chromosomal *E. uredovora* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CATTGAGAAGCTTATGTGCACCG-3', 10 pmol of the primer 5'-CTCCGGGGTCGACATGGCGC-3', 40 ng of chromosomal DNA of *E. uredovora*, 8 U of Taq DNA polymerase (Eurogentec), 20 nmol of dNTPs, Taq Extender (Stratagene) in a total volume of 100 µl 1x Taq Extender buffer (Stratagene).

The mixture is denaturated for 3 min at 94 °C. Then 40 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 300 sec at 72 °C followed. After further incubation for 20 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden, Germany).

1.0 µg of the vector pBluescript SKII$^-$ (Stratagene) and 2.0 µg of the purified PCR product are digested with *Hind*III and *Sal*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 40 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScaro34. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScaro34 is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScaro34 is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb D90087).

The *E. uredovora* ORF *crtE* (accession no. gb D90087) from base pair (bp) position 175 to 1148 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CCGCATCTTTCCAATTGCCG-3', 10 pmol of the primer 5'-ATGCAGCAAGCTTAACTGACGGC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA

polymerase (Eurogentec, Seraing, Belgium) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden, Germany).

1.5 µg of the vector pBScaro34 (see above) is digested with *Eco*RI and *Hind*III and 0.6 µg of the purified PCR product are digested with *Mfe*I and *Hind*III in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 16 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScaro14. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScaro14 is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScaro14 is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb D90087). The DNA sequence of the plasmid pBScaro14 is shown in Annex E.

5 µg of the vector pBScaro14 (see above) is digested with *Bam*HI and *Sal*I. The restriction mixture is prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. The restriction mixture is separated on a agarose gel and the fragments of 2237 and 2341 bp size are purified with the gel extraction kit from Qiagen.

3 µg of the vector pACYC184 (see above) is digested with *Bam*HI and *Sal*I. The restriction

mixture is prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. The restriction mixture is separated on a agarose gel and the fragment of 3968 bp size is purified with the gel extraction kit from Qiagen.

30 ng of the purified vector DNA and each 25 ng of the purified 2237 and 2341 bp fragments are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYCcaro14. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYCcaro14 is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYCcaro14 is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb D90087). The DNA sequence of the plasmid pACYCcaro14 is shown in Annex F.

Example 8

Enzymatic preparation of [U-$^{13}C_5$]1-deoxy-D-xylulose 5-phosphate

A reaction mixture containing 960 mg of [U-$^{13}C_6$]glucose (5.1 mmol), 6.1 g of ATP (10.2 mmol), 337 mg of thiamine pyrophosphate, 1.14 g of [2,3-$^{13}C_2$]pyruvate (10.2 mmol), 10 mM $MgCl_2$, 5 mM dithiothreitol in 150 mM Tris hydrochloride, pH 8.0 is prepared. 410 Units of triose phosphate isomerase (from rabbit muscle, Type III-S, E. C. 5.3.1.1., Sigma), 100 U hexokinase (from Bakers Yeast, Type VI, E. C. 2.7.1.1, Sigma), 100 U phosphoglucose isomerase (from Bakers Yeast, Type III, E. C. 5.3.1.9, Sigma), 100 U phosphofructokinase (from *Bacillus stearothermophilus*, Type VII, E. C. 2.7.1.11, Sigma), 50 U aldolase (from rabbit muscle, E. C. 4.1.2.13, Sigma) and 12 U of recombinant DXP synthase from *B. subtilis* are added to a final volume of 315 ml. The reaction mixture is incubated at 37 °C overnight and during incubation the pH is hold at a constant value of 8.0. The reaction is monitored by ^{13}C NMR spectroscopy.

Example 9

Enzymatic preparation of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose 5-phosphate

A solution containing 150 mM Tris hydrochloride, 10 mM $MgCl_2$, 1.0 g of [U-$^{13}C_6$]glucose (5.4 mmol), 0.23 g (1.5 mmol) of dithiothreitol, 0.3 g (0.7 mmol) of thiamine pyrophosphate, 0.1 g (0.2 mmol) of ATP (disodium salt), and 2.2 g (11mmol) of phosphoenol pyruvate (potassium salt) is adjusted to pH 8.0 by the addition of 8 M sodium hydroxide. 403 U (2.8 mg) of pyruvate kinase (from rabbit muscle, E. C. 2.7.1.40), 410 Units of triose phosphate isomerase (from rabbit muscle, Type III-S, E. C. 5.3.1.1., Sigma), 100 U hexokinase (from Bakers Yeast, Type VI, E. C. 2.7.1.1, Sigma), 100 U phosphoglucose isomerase (from Bakers Yeast, Type III, E. C. 5.3.1.9, Sigma), 100 U phosphofructokinase (from *Bacillus stearothermophilus*, Type VII, E. C. 2.7.1.11, Sigma), 50 U aldolase (from rabbit muscle, E. C. 4.1.2.13, Sigma) and 12 U recombinant DXP synthase from *B. subtilis* are added to a final volume of 300 ml. The reaction mixture is incubated at 37 °C for overnight.

Example 10

Enzymatic preparation of 1-deoxy-D-xylulose

The pH value of the reaction mixture obtained in example 8 or 9 is adjusted to 9.5. Magnesium chloride is added to a concentration of 30 mM. 50 mg (950 Units) of alkaline phosphatase from bovine intestinal mucosa (Sigma, E. C. 3.1.3.1) are added and the reaction mixture is incubated for 16 h. The conversion is monitored by ^{13}C-NMR spectroscopy. The pH is adjusted to a value of 7.0 and the solution is centrifuged at 14,000 upm for 5 minutes. Starting from labelled glucose (examples 8 or 9) the overall yield of 1-deoxy-D-xylulose is approximately 50 %.

The supernatant or the lyophilised supernatant is used in incorporation experiments (see examples 11 to 17).

Example 11

Incorporation experiment with recombinant *Escherichia coli* XL1-pBSxylB using [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose

0.2 litre of Luria Bertani (LB) medium containing 36 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harbouring the plasmid pBSxylB (see example 1). The cells are grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.6 the culture is induced with 2 mM IPTG. Two hours after induction with IPTG, 50 ml (0.9 mmol) of

crude [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose (pH 7.0) (see examples 9 and 10), are added. Aliquots of 25 ml are taken at time intervals of 30 minutes and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. ^{13}C NMR spectra of the supernatant are recorded directly, without further purification, with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). The NMR analysis is based on published signal assignments (Wungsintaweekul *et al.*, 2001).

30 min after the addition of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose, the formation of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose 5-phosphate can be observed. The maximum yield of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose 5-phosphate is observed 3-5 h after addition of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose to the medium. The ^{13}C NMR signals reveal a mixture of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose and [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose-5-phosphate at a molar ratio of approximately 1 : 9. The intracellular concentration of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose 5-phosphate is estimated as 20 mM by quantitative NMR spectroscopy.

Example 12

Incorporation experiment with recombinant *Escherichia coli* XL1-pBSxylBdxr using [U-$^{13}C_5$]1-deoxy-D-xylulose

0.12 litre of Luria Bertani (LB) medium containing 22 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harbouring plasmid pBSxylBdxr (see example 2). The cells are grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.6 the culture is induced with 2 mM IPTG. Two hours after induction with IPTG, ca. 1.0 mmol of crude [U-$^{13}C_5$]1-deoxy-D-xylulose (pH 7.0) (see examples 8 and 10) are added. Aliquots of 25 ml are taken in time intervals of 1 h and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the supernatant are recorded directly, without further purification, with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany).

HMQC and HMQC-TOCSY experiments reveal ^{1}H-^{13}C and ^{1}H-^{1}H spin systems (Table 1) of [U-$^{13}C_5$]2*C*-methyl-D-erythritol 4-phosphate, [U-$^{13}C_5$]2*C*-methyl-D-erythritol and [1,2,2´,3,4-$^{13}C_5$]4-diphosphocytidyl-2*C*-methyl-D-erythritol at a molar ratio of approximately 6.6 : 7 : 1, respectively. The intracellular concentration of [U-$^{13}C_5$]2*C*-methyl-D-erythritol 4-phosphate is estimated as 10 mM by quantitative NMR spectroscopy.

The NMR data summarized in Table 1 are identical with published NMR data of the authentic compounds (Takahashi et al., 1998; Rohdich et al., 1999).

Table 1. NMR data of ^{13}C-labeled products in cell extracts of *E. coli* XL1-pBSxylBdxr after feeding of [U-$^{13}C_5$]1-deoxy-D-xylulose

	Chemical shifts, ppm						
Position	1	1*	2	2-Methyl	3	4	4*
[U-$^{13}C_5$]2*C*-methyl-D-erythritol 4-phosphate							
^{13}C	66.1		n.d.	18.1	73.4	648	
^{1}H	3.25	3.36		0.93	3.56	3.62	3.81
[U-$^{13}C_5$]2*C*-methyl-D-erythritol							
^{13}C	66.6		n.d.	18.0	74.6	616	
^{1}H	3.26	3.34		0.9	3.44	3.36	3.61
[1,2,2´,3,4-$^{13}C_5$]4-diphosphocytidyl-2*C*-methyl-D-erythritol							
^{13}C	66.8		n.d.	18.0	73.0	667	
^{1}H	3.4	3.55		0.9	3.6	3.74	4

Example 13

Incorporation experiment with recombinant *Escherichia coli* XL1-pBSxylBdxrispDF using [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose

0.1 litre of Luria Bertani (LB) medium containing 18 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harbouring the plasmid pBSxylBdxrispDF (see example 4). The cells are grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.5, the culture is induced with 2 mM IPTG. Two hours after induction with IPTG, ca. 1.0 mmol of crude [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose (see examples 9 and 10) are added. After three

hours, cells were harvested and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 1.5 ml of 20 mM NaF in D_2O, cooled on ice and sonified 3 x 15 sec. with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the supernatant are recorded directly, without further purification, with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany).

HMQC and HMQC-TOCSY experiments reveal ^1H-^{13}C and ^1H-^1H spin systems of [1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol 4-phosphate, [1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol and [1,3,4-$^{13}C_5$]4-diphosphocytidyl-2*C*-methyl-D-erythritol (Table 1). The molar ratios of [1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol 4-phosphate, [1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol and [1,3,4-$^{13}C_5$]4-diphosphocytidyl-2*C*-methyl-D-erythritol are 1 : 0.6 : 0.9, respectively.

This result indicates that the intracellular amount of CTP required for the synthesis of 4-diphosphocytidyl-2*C*-methyl-D-erythritol is limiting. Therefore, a modified fermentation protocol was developed (see example 14).

Example 14

Incorporation experiment with recombinant *Escherichia coli* XL1-pBSxylBdxrispDF using [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose

0.1 litre of Luria Bertani (LB) medium containing 18 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harbouring plasmid pBSxylBispDF (see example 4). The cells are grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.5, the culture is induced with 2 mM IPTG. Two hours after induction with IPTG, 10 mg (0.041 mmol) of cytidine and 5 ml of 1 M $NaKHPO_4$, pH 7.2, and ca. 1 mmol of crude [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose (see examples 9 and 10) are added. After three hours, the cells are harvested and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the supernatant are recorded directly, without further purification, with a Bruker AVANCE DRX 500 spectrometer. HMQC and HMQC-TOCSY experiments reveal ^1H-^{13}C and ^1H-^1H spin systems (Table 1) of

[1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol 4-phosphate, [1,3,4-$^{13}C_3$]2*C*-methyl-D-erythritol and [1,3,4-$^{13}C_3$]4-diphosphocytidyl-2*C*-methyl-D-erythritol at a molar ratio of approximately 1 : 3.4 : 4.2, respectively. The relative amount of [1,3,4-$^{13}C_3$]4-diphosphocytidyl-2*C*-methyl-D-erythritol is increased by a factor of 2 as compared to the relative amount in example 13. The intracellular concentration of [1,3,4-$^{13}C_3$]4-diphosphocytidyl-2*C*-methyl-D-erythritol is estimated as 10 mM by quantitative NMR spectroscopy. The relative high amount of 2*C*-methyl-D-erythritol indicates that unspecific phosphatases convert intermediary formed 2*C*-methyl-D-erythritol 4-phosphate into 2*C*-methyl-D-erythritol. Therefore, a modified fermentation protocol was developed to supply the cells with sufficient amounts of organic phosphates and in order to suppress the activity of phosphatases (see examples 15 to 17).

Example 15

Incorporation experiment with recombinant *Escherichia coli* XL1-pBScyclo using [U-$^{13}C_5$]1-deoxy-D-xylulose

0.2 litre of Luria Bertani (LB) medium containing 36 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harbouring the plasmid pBScyclo (see example 5). The cells are grown in a shaking culture at 37 °C. At an optical density (600 nm) of 1.3, the culture is induced with 2 mM IPTG. Two hours after induction with IPTG, 30 mg (0.12 mmol) of cytidine, 300 mg (0.95 mmol) of DL-α-glycerol 3-phosphate and 10 ml of 1 M $NaKHPO_4$, pH 7.2, are added. After 30 min, ca. 1 mmol of [U-$^{13}C_5$]1-deoxy-D-xylulose (see example 8 and 10) are added. Aliquots of 25 ml are taken at time intervals of 1 h and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the cell free extract are recorded directly, without further purification, with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). ^{13}C NMR spectra, as well as HMQC and HMQC-TOCSY spectra established [U-$^{13}C_5$]2*C*-methyl-D-erythritol 2,4-cyclodiphosphate (Herz *et al.*, 2000) as the only product. Formation of [U-$^{13}C_5$]2*C*-methyl-D-erythritol 2,4-cyclodiphosphate can be observed 30 min after addition of [U-$^{13}C_5$]1-deoxy-D-xylulose, whereas the maximum yield is observed 5 h after addition of [U-$^{13}C_5$]1-deoxy-D-xylulose. The intracellular concentration of [U-$^{13}C_5$]2*C*-methyl-D-

erythritol 2,4-cyclodiphosphate is estimated as 20 mM by quantitative NMR spectroscopy. The formation of any other isotope-labelled products, such as [U-$^{13}C_5$]2*C*-methyl-erythritol is completely suppressed.

Example 16

Incorporation experiment with recombinant *Escherichia coli* XL1-pBScyclo-pACYCgcpE using [2-^{14}C]- and [U-$^{13}C_5$]1-deoxy-D-xylulose

0.2 litre of Terrific Broth (TB) medium containing 36 mg of ampicillin and 2.5 mg of chloramphenicol are inoculated with the *E. coli* strain XL1-Blue harbouring the plasmids pBScyclo and pACYCgcpE (see example 5 to 6). The cells are grown in a shaking culture at 37 °C overnight. At an optical density (600 nm) of 4.8 to 5.0, 30 mg (0.1 mmol) of cytidine, 300 mg (0.94 mmol) of DL-α-glycerol 3-phosphate and 10 ml of 1 M $NaKHPO_4$, pH 7.2, are added. After 30 minutes, a mixture of 2.6 µmol [2-^{14}C]1-deoxy-D-xylulose (15 µCi µmol^{-1}) (Wungsintaweekul et al., 2001) and 1 ml of crude [U-$^{13}C_5$]1-deoxy-D-xylulose (0.02 mmol) (see examples 8 and 10) are added. After 1.5 h, cells are harvested and centrifuged for 10 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in a mixture of 20 mM NaF (2 ml) and methanol (2 ml), cooled on ice and sonified 3 x 15 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4. The suspension is centrifuged at 15,000 rpm for 15 min. The radioactivity of the supernatant is measured by scintillation counting (Beckmann, LS 7800). 10 % of the radioactivity initially added as ^{14}C labelled 1-deoxy-D-xylulose is detected in the supernatant. Aliquots are analysed by TLC and HPLC, as described in example 19, and the products are purified as described in example 20. On basis of these data, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate were identified as products at a molar ratio of 7 : 3 (see also examples 17 and 18).

Example 17

Incorporation experiment with recombinant *Escherichia coli* XL1-pBScyclo-pACYCgcpE using [U-$^{13}C_5$]- or [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose

0.2 litre of Terrific Broth (TB) medium containing 36 mg of ampicillin and 2.5 mg of

chloramphenicol are inoculated with the *E. coli* strain XL1-Blue harbouring the plasmids pBScyclo and pACYCgcpE. The cells are grown in a shaking culture at 37 °C for overnight. At an optical density (600 nm) of 4.8 - 5.0, 30 mg (0.1 mmol) of cytidine, 300 mg (0.93 mmol) of DL-α-glycerol 3-phosphate and 10 ml of 1 M $NaKHPO_4$, pH 7.2, are added. After 30 minutes, 3 ml of crude [3,4,5-$^{13}C_3$]- or [U-$^{13}C_5$]1-deoxy-D-xylulose (0.05 mmol) (see examples 8, 9, and 10) are added. Aliquots of 25 ml are taken at time intervals of 1 h and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O or in 700 µl of a mixture of methanol and D_2O (6:4; v/v) containing 10 mM NaF, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the cell free extracts are recorded directly with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). In order to avoid degradation during work-up, the structures of the products are determined by NMR spectroscopy without further purification (see example 18).

Example 18

Structure determination of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate

The ^{1}H-decoupled ^{13}C NMR spectrum using [U-$^{13}C_5$]1-deoxy-D-xylulose as starting material displays 5 ^{13}C-^{13}C coupled signals belonging to 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate (Herz et al., 2000) and 5 ^{13}C-^{13}C coupled signals at 14.7, 64.5, 68.6, 122.7 and 139.5 ppm (Table 2) belonging to an unknown metabolite. The chemical shifts of the unknown metabolite suggest a double bond motif (signals at 122.7 and 139.5 ppm), a methyl group (signal at 14.7 ppm), and two carbon atoms (signals at 64.5 and 68.6 ppm) connected to OR (R=unknown). The three signals accounting for carbon atoms with sp^3 hybridisation (14.7, 64.5 and 68.5 ppm) show ^{13}C-^{13}C coupling to one adjacent ^{13}C atom with coupling constants of 40 - 50 Hz (Table 2). The signal at 122.7 ppm shows ^{13}C couplings to two adjacent ^{13}C neighbours (coupling constants, 74 and 50 Hz), whereas the signal at 141.5 ppm shows ^{13}C couplings to three neighboured ^{13}C atoms (coupling constants, 74, 43 and 43 Hz). In conjunction with the chemical shift topology, this coupling signature is indicative for a 2-methyl-2-butenyl skeleton. HMQC and HMQC-TOCSY experiments reveal the ^{1}H NMR chemical shifts (Table 2), as well as ^{13}C-^{1}H and ^{1}H-^{1}H spin systems (Table 3). More specifically, the ^{13}C NMR signal at 122.7

ppm correlates to a ^{1}H NMR signal at 5.6 ppm which is in the typical chemical shift range for H-atoms attached to CC double bonds, whereas the signal at 139.5 ppm gives no ^{13}C-^{1}H correlations. The signals at 64.5 and 68.6 ppm give ^{13}C-^{1}H correlations to ^{1}H-signals at 4.5 and 3.9 ppm, respectively. The methyl signal at 14.7 ppm correlates to a proton signal at 1.5 ppm. In connection with ^{13}C-^{13}C coupling patterns (Table 2), as well as with ^{1}H-^{13}C long range correlations (HMBC experiment, Table 3), these data establish a 1,4-dihydroxy-2-methyl-2-butenyl system.

Starting from [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose as feeding material three signals at 64.5, 68.6 and 122.7 ppm accounting for atoms 4, 1 and 3, respectively, of the new product are observed. It can be concluded that the carbon atoms at 1, 3 and 4 of the new product are biogenetically equivalent to the carbon atoms 3, 4 and 5 of [3,4,5-$^{13}C_3$]1-deoxy-D-xylulose 5-phosphate. This coupling topology is similar to the coupling pattern of 2C-methyl-D-erythritol 4-phosphate (see example 13) confirming that the new compound is derived via 2C-methyl-D-erythritol 4-phosphate.

The C-4 and C-3 ^{13}C NMR signals at 64.5 and 122.7 ppm show ^{13}C-^{31}P coupling of 5.5 and 8.0 Hz, respectively. These couplings indicate the presence of a phosphate or pyrophosphate group at position 4 of the 2-methyl-2-butenyl skeleton.

In line with this observation, the ^{1}H-decoupled ^{31}P NMR spectrum of the product displays a doublet at –9.2 (^{31}P-^{31}P coupling constant, 20.9 Hz) and a double-double-doublet at –10.6 ppm (^{31}P-^{13}C coupling constants, 5.8 and 7.4 Hz, ^{31}P-^{31}P coupling constant, 20.9 Hz). Without ^{1}H-decoupling, the ^{31}P NMR signal at –10.6 ppm is broadened whereas the signal at –9.2 ppm is not affected by ^{1}H coupling. The chemical shifts as well as the observed coupling pattern confirm the presence of a free diphosphate moiety at position 4.

In summary, all these data establish the structure as 1-hydroxy-2-methyl-2-butenyl 4-diphosphate.

Table 2. NMR data of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate

Position	Chemical shifts, ppm			Coupling constants, Hz		
	$^{1}H^{a}$	$^{13}C^{b}$	$^{31}P^{c}$	J_{PC}	J_{PP}	J_{CC}
1	3.91	68.6[d,e]				43.0[e], 5.5[d], 3.5[d]
2		139.5[d,e]				74.3[e], 43.3[e], 43.3[e]
2-Methyl	1.51	14.7[e]				42.2[e], 4.0[e], 4.0[e]
3	5.57	122.7[e]		8.0[d]		73.9[e], 49.8[d], 4.0[d]
4	4.46	64.5[d,e]		5.5[d]		49.3[d], 5.5[d]
P_{β}			-9.2		20.9	
P_{α}			-10.6	5.8[d], 7.4[d]	20.9	

[a]referenced to external trimethylsilylpropane sulfonate.

[b]referenced to external trimethylsilylpropane sulfonate.

[c]referenced to external 85 % orthophosphoric acid.

[d]observed with [1,3,4-$^{13}C_3$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate

[e]observed with [U-$^{13}C_5$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate

Table 3. Correlation pattern of [1,3,4-$^{13}C_3$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate and of [U-$^{13}C_5$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate

NMR Correlation pattern			
Position	HMQC	HMQC-TOCSY	HMBC
1	1[a,b]	1[a,b]	2-methyl[a], 2[a]
2			
2-methyl	2-methyl[b]	2-methyl[b]	
3	3[a,b]	3[a,b], 4[a,b]	2-methyl[a], 1[a]
4	4[a,b]	4[a,b], 3[a,b]	

[a]observed with [1,3,4-$^{13}C_3$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate

[b]observed with [U-$^{13}C_5$]1-hydroxy-2-methyl-2-butenyl 4-diphosphate

Example 19

Detection of phosphorylated metabolites of the mevalonate-independent pathway

Method A) By a TLC method

Aliquots (10 µl) of the cell-free extracts from recombinant cells prepared as described above (see example 16) are spotted on a Polygram® SIL NH-R thin layer plate (Macherey-Nagel, Düren, Germany). The TLC plate is then developed in a solvent system of n-propanol: ethyl acetate: water; 6:1:3 (v/v/v). The running time is about 4 h. The radio chromatogram is monitored and evaluated by a Phosphor Imager (Storm 860, Molecular Dynamics, USA). The R_f-values of the compounds under study are shown in Table 4.

Table 4: R_f-values of precursors and intermediates of the mevalonate-independent terpenoid pathway

Chemical compound	R_f-value
1-deoxy-D-xylulose	0.80
1-deoxy-D-xylulose 5-phosphate	0.5
2*C*-methyl-D-erythritol 4-phosphate	0.42
4-diphosphocytidyl-2*C*-methyl-D-erythritol	0.33
4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate	0.27
2*C*-methyl-D-erythritol 2,4-cyclodiphosphate	0.47
1-hydroxy-2-methyl-2-butenyl 4-diphosphate	0.17

Method B) By a HPLC method

Aliquots (100 µl) of the cell-free extracts from recombinant cells prepared as described above (see example 16), are analyzed by HPLC using a column of Multospher 120 RP 18-AQ-5 (4.6 x 250 mm, particle size 5 µm, CS-Chromatographic Service GmbH, Langerwehe, Germany) that has been equilibrated for 15 min with 10 mM tetrabutylammonium hydrogensulfate (TBAS), pH 6.0, at a flow rate of 0.75 ml min^{-1}. After injection of the sample, the column is developed for 20 min with 10 mM TBAS, then for 60 min with a linear gradient of 0 - 42 % (v/v)

methanol in 10 mM TBAS. The effluent is monitored by a continuous-flow radio detector (Beta-RAM, Biostep GmbH, Jahnsdorf, Germany). The retention volumes of the compounds under study are shown in Table 5.

Table 5: Retention volumes of precursors and intermediates of the mevalonate-independent terpenoid pathway

Chemical compound	Retention volume [ml]
1-deoxy-D-xylulose	6.0
1-deoxy-D-xylulose 5-phosphate	15
2*C*-methyl-D-erythritol 4-phosphate	13.5
4-diphosphocytidyl-2*C*-methyl-D-erythritol	30.8
4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate	41.3
2*C*-methyl-D-erythritol 2,4-cyclodiphosphate	31.5
1-hydroxy-2-methyl-2-butenyl 4-diphosphate	42.8

Example 20

Purification of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate

The crude cell free extract obtained from the feeding experiment with recombinant *Escherichia coli* XL1-pBScyclo-pACYCgcpE using [2-^{14}C]- and [U-$^{13}C_5$]1-deoxy-D-xylulose (see example 16) is lyophilized. The residue is dissolved in 600 µl of water and centrifuged for 10 min at 14,000 ppm. Aliquots of 90 µl are applied on a column of Nucleosil 10 SB (4.6 x 250 mm, Macherey & Nagel, Düren, Germany) which is developed with a linear gradient of 0.1 - 0.25 M ammonium formate in 70 ml at a flow rate of 2 ml min^{-1}. The retention volumes for 2*C*-methyl-D-erythritol-2,4-cyclodiphosphate and 1-hydroxy-2-methyl-2-butenyl 4-diphosphate are 25 and 44 ml, respectively. Fractions are collected and lyophilized. NMR data of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate are identical with the data shown in example 18, Table 2.

Example 21

Construction of a vector carrying the *xylB*, *dxr*, *ispD*, *ispE*, *ispF* and *ispG* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase, CDP-

ME synthase, CDP-ME kinase cMEPP synthase and 1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase

The *E. coli* ORF *ispG* (accession no. gb AE000338) from base pair (bp) position 372 to 1204 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GCGGGAGACCGCGGGAGGAGAAATTAACCATGCATAACCAGGCTCCAATTCG-3', 10 pmol of the primer 5'-CGCTTCCCAGCGGCCGCTTATTTTTCAACCTGCTGAACG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.4 µg of the vector pBScyclo (Example 5) and 0.8 µg of the purified PCR product are digested with *Sac*II and *Not*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 18 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScyclogcpE. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScyclogcpE is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScyclogcpE is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000338). The DNA sequence of

the vector construct pBScyclogcpE is shown in Annex H.

Example 22

Construction of a vector carrying the *xylB*, *dxr*, *ispD*, *ispE*, *ispF*, *ispG* and *lytB* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase, CDP-ME synthase, CDP-ME kinase, cMEPP synthase, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase and LytB

The *E. coli* ORF *lytB* (accession no. gb AE005179) from base pair (bp) position 7504 to 8454 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-AAATCGGAGCTCGAGGAGAAATTAACCATGCAGATCCTGTTGGCC-3', 10 pmol of the primer 5'-GCTGCTCCGCGGTTAATCGACTTCACGAATATCG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.3 µg of the vector pBScyclogcpE (Example 21) and 0.7 µg of the purified PCR product are digested with *Sac*I and *Sac*II in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 16 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScyclogcpElytB. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScyclogcpElytB is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScyclogcpElytB is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE005179).

Example 23

Incorporation experiment with recombinant *Escherichia coli* XL1-pBScyclogcpE using [U-$^{13}C_5$]1-deoxy-D-xylulose

0.1 litre of Terrific Broth (TB) medium containing 18 mg of ampicillin are inoculated with *E. coli* strain XI1-Blue harbouring the plasmid pBScyclogcpE. The cells are grown in a shaking culture at 37 °C overnight. At an optical density (600 nm) of 4.8 – 5.0, 30 mg (0.1 mmol) of cytidine are added. A solution containing 1.2 g of lithium lactate (12.5 mmol), 6 ml of crude [U-$^{13}C_5$]1-deoxy-D-xylulose (0.05 mmol) (see examples 8, 9 and 10) in 0.1 M Tris hydrochloride (pH=7.5) at a final volume of 30 ml are added continously within 2 hours. Aliquots of 25 ml are taken at time intervals of 1 h and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of 20 mM NaF in D_2O or in 700 µl of a mixture of methanol and D_2O (6:4, v/v) containing 10 mM NaF, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the cell free extracts are recorded directly with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). In order to avoid degradation during work-up, the structures of the products are determined by NMR spectroscopy without further purification.

The relative amount of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate to 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate could be raised by a factor of approximately 2-3 by the addition of lithium lactate to the medium.

Example 24

Preparation of (*E*)-1-hydroxy-2-methyl-2-butenyl diphosphate triammonium salt (8)

General. Chemicals are obtained from Acros Organics (Fisher Scientific GmbH, Schwerte, Germany), SIGMA-ALDRICH (Deisenhofen, Germany), MERCK (Darmstadt, Germany) and used without further purification. Solvents are used distilled and/or dried. Chromatography is performed on silica gel 60 (230-400 mesh, Fluka Riedel-de Haen, Taufkirchen, Germany), DOWEX 50 WX8 (200-400 mesh, SERVA, Heidelberg, Germany), and Cellulose (Avicel, Cellulose mikrokristallin, Merck, Darmstadt, Germany). TLC is performed on silica gel 60 F_{254} plastic sheets (MERCK) or cellulose F plastic sheets (MERCK), detection by anisaldeyde solution (anisaldehyde:H_2SO_4:HAc 0.5:1:50 v/v/v). NMR-spectra are recorded on BRUKER AMX 400, DRX 500, and AC 250 spectrometer at room temperature.

Acetonyl tetrahydropyranyl ether (12) (Hagiwara *et al.*, 1984).
A mixture of 339 mg (1.35 mmol) of pyridinium-toluene-4-sulfonate, 9.35 ml (10.0 g, 0.135 mol) of hydroxyacetone, and 24.7 ml (22.7 g, 0.270 mol) of 3,4-dihydro-2*H*-pyran is stirred at room temperature for 2.5 h. Residual 3,4-dihydro-2*H*-pyran is removed under reduced pressure. The crude mixture is purified by FC on silicagel (hexanes/acetone 4:1, 6.5 × 20 cm) to yield 18.7 g (0.118 mol, 88%) of a colorless liquid.
^{1}H NMR ($CDCl_3$, 500 MHz) δ 4.62 (t, *J* = 3.6 Hz, 1H), 4.22 (d, *J* = 17.3 Hz, 1H), 4.09 (d, *J* = 17.3 Hz, 1H), 3.83-3.79 (m, 1H), 3.51-3.47 (m, 1H), 2.15 (s, 3H), 1.87-1.49 (m, 6H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 206.7, 98.7, 72.3, 62.3, 30.2, 26.5, 25.2, 19.2; MS (CI, isobutane) *m/z* 159 $[M + 1]^+$.

(E,Z)-Ethyl-2-methyl-1-tetrahydropyranyloxy-but-2-enoate (13) (Watanabe *et al.*, 1996).
33.0 g (94.8 mmol) of (ethoxycarbonylmethylen)-triphenylphosphorane are dissolved in 500 ml of dry toluene under nitrogen atmosphere at room temperature. Then, 10.0 g (63.2 mmol) of acetonyl tetrahydropyranyl ether **12** are added and the mixture is heated to reflux. After 39 h at this temperature the solvent is evaporated *under reduced pressure* to yield an orange oil. Major amounts of triphenylphosphinoxide are precipitated by the addition of 100 ml hexanes/acetone 9:1. After filtration the filtrate is concentrated and another 100 ml of hexanes/acetone 9:1 are added. The solid is filtered off and the solvent removed to yield 18 g of an orange oil that is purified by FC on silicagel (hexanes/acetone 9:1, 6.5 × 28 cm) to yield 12.9 g (56.5 mmol, 89%) of a mixture of **(E)-13/(Z)-13** = 5:1.
(E)-(13). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.96 (q, *J* = 1.4 Hz, 1H), 4.62 (t, *J* = 3.5 Hz, 1H), 4.20 (dd, *J* = 15.5 Hz, 1.3 Hz, 1H), 4.14 (q, *J* = 7.1 Hz, 2H), 3.93 (dd, *J* = 15.6, 1.3 Hz, 1H), 3.84-

3.79 (m, 1H), 3.52-3.48 (m, 1H), 2.08 (d, *J* = 1.4 Hz, 3H), 1.88-1.50 (m, 6H), 1.26 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 166.8, 154.7, 114.5, 98.0, 70.6, 62.0, 59.7, 30.3, 25.3, 19.1, 15.9, 14.3; MS (CI, isobutane) *m/z* 229 $[M + 1]^+$.

(Z)-(13). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.71 (q, *J* = 1.4 Hz, 1H), 4.60 (t, *J* = 3.6 Hz, 1H), 4.20 (dd, *J* = 15.5 Hz, 1.3 Hz, 1H), 4.11 (q, *J* = 7.1 Hz, 2H), 3.93 (dd, *J* = 15.6, 1.3 Hz, 1H), 3.84-3.79 (m, 1H), 3.52-3.48 (m, 1H), 1.97 (d, *J* = 1.4 Hz, 3H), 1.88-1.50 (m, 6H), 1.24 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 165.9, 156.8, 116.9, 98.7, 66.5, 62.3, 59.8, 30.6, 25.3, 21.9, 19.5, 14.3; MS (CI, isobutane) *m/z* 229 $[M + 1]^+$.

(E,Z)-2-Methyl 1-tetrahydropyranyloxy-but-2-ene-4-ol (14) (Watanabe *et al.*, 1996).

A solution of ester **13** (8.73 g, 38.2 mmol) in 100 ml of dry CH_2Cl_2 is cooled to -78 °C. Then, 91.8 ml (91.8 mmol) of 1.0 M DIBAH in hexanes are added slowly under an atmosphere of nitrogen. The resulting solution is stirred for 3 h at -78 °C before the reaction is quenched by the addtion of 1.5 ml of 1 M NaOH. After warming to room temperature the solvent is removed *under reduced pressure*. The resulting gummy residue is widely dissolved by adding twice 100 ml of MeOH. The resulting mixture is passed through a column of SiO_2, evaporated from the solvent and then loaded on a column of SiO_2/Na_2SO_4 that is purged with 1400 ml of MeOH. Evaporation of the solvent gives 9.5 g of a colorless liquid that is purified by FC on silica gel (hexanes/acetone 1:3, 6.5 × 16 cm) to yield 6.98 g (37.4 mmol, 98%) of a colorless liquid **(E)-14/(Z)-14** = 6:1.

(E)-(14). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.68 (tq, *J* = 6.6, 1.3 Hz, 1H), 4.60 (t, *J* = 3.6 Hz, 1H), 4.20 (d, *J* = 6.8 Hz, 2H), 4.12 (d, *J* = 12.0 Hz, 1H), 3.87-3.82 (m, 1H), 3.85 (d, *J* = 12.5 Hz, 1H), 3.52-3.48 (m, 1H), 1.86-1.48 (m, 6H), 1.69 (s, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 135.7, 125.4, 97.8, 71.9, 62.1, 59.1, 30.5, 25.4, 19.4, 14.1; MS (CI, isobutane) *m/z* 169 $[M - H_2O + 1]^+$.

(Z)-(14). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.64 (tq, *J* = 6.6, 1.3 Hz, 1H), 4.63 (t, *J* = 3.3 Hz, 1H), 4.20 (d, *J* = 6.8 Hz, 2H), 4.15 (d, *J* = 11.8 Hz, 1H), 3.87-3.82 (m, 1H), 3.83 (d, *J* = 11.3 Hz, 1H), 3.52-3.48 (m, 1H), 1.86-1.48 (m, 6H), 1.79 (s, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 136.2, 128.6, 96.6, 65.1, 61.8, 58.1, 30.3, 25.3, 21.9, 19.0; MS (CI, isobutane) *m/z* 169 $[M-H_2O+1]^+$.

(E,Z)-4-Chloro-2-methyl 1-tetrahydropyranyloxy-but-2-en (15) (Hwang *et al.*, 1984).

To a solution of alcohol **14** (1.00 g, 5.37 mmol) in 10 ml of dry CH_2Cl_2 are added 918 mg (7.52 mmol) of DMAP in 10 ml of dry CH_2Cl_2 and 1.23 g (6.44 mmol) of *p*-TsCl in 10 ml of dry

CH_2Cl_2. The resulting solution is stirred at room temperature for 1 h. After evaporation of the solvent *under reduced pressure* the residue is purified by FC on silica gel (CH_2Cl_2, 5 × 20 cm) to obtain 693 mg (3.39 mmol, 63%) of a colorless liquid **(E)-15/(Z)-15** = 6:1.

(E)-(15). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.77 (tq, *J* = 8.0, 1.5 Hz, 1H), 4.64 (t, *J* = 3.6 Hz, 1H), 4.18 (d, *J* = 12.8 Hz, 1H), 4.15 (d, *J* = 8.0 Hz, 2H), 3.92 (d, *J* = 12.8 Hz, 1H), 3.90-3.86 (m, 1H), 3.59-3.52 (m, 1H), 1.92-1.52 (m, 6H), 1.77 (s, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 138.6, 121.7, 97.8, 71.3, 62.1, 40.2, 30.5, 25.4, 19.3, 13.9; MS (CI, isobutane) *m/z* 205 $[M + 1]^+$.

(Z)-(15). ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.65 (t, *J* = 8.1 Hz, 1H), 4.61 (t, *J* = 3.6 Hz, 1H), 4.18 (d, *J* = 12.8 Hz, 1H), 4.15 (d, *J* = 8.0 Hz, 2H), 3.92 (d, *J* = 12.8 Hz, 1H), 3.90-3.86 (m, 1H), 3.59-3.52 (m, 1H), 1.92-1.52 (m, 6H), 1.86 (s, 3H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 138.3, 124.6, 97.5, 64.7, 62.2, 40.1, 30.5, 25.4, 21.8, 19.4; MS (CI, isobutane) *m/z* 205 $[M + 1]^+$.

(E,Z)-2-Methyl 1-tetrahydropyranyloxy-but-2-enyl diphosphate triammonium salt (16) (Davisson *et al.*, 1986).

To a solution of chloride **15** (260 mg, 1.27 mmol) in 1.3 ml of MeCN a solution of 1.38 g (1.52 mmol) tris(tetra-*n*-butylammonium) hydrogen pyrophosphate in 3.0 ml of MeCN is added slowly at room temperature, obtaining an orange-red solution. The reaction is followed by ^{1}H-NMR, taking advantage of the up field shift of the multiplet of H-3. After 2 h the reaction is finished and the solvent removed *under reduced pressure*. The orange oil is dissolved in 2.5 ml of H_2O and passed through a column of DOWEX 50 WX8 (2.5 × 3 cm) cation-exchange resin (NH_4^+ form) that has been equilibrated with two column volumes (40 ml) of 25 mM NH_4HCO_3. The column is eluted with 60 ml of 25 mM NH_4HCO_3. The resulting solution is lyophilized, dissolved in 5 ml of isopropanol/100 mM NH_4HCO_3 1:1 and loaded on a cellulose column (2 × 18 cm) that is eluted by isopropanol/100 mM NH_4HCO_3 1:1. The effluent is lyophilized obtaining 495 mg (1.25 mmol, 98%) of **(E)-16/(Z)-16** = 6:1 as a white solid.

(E)-(16). ^{1}H NMR (D_2O, 500 MHz) δ 5.52 (tq, *J* = 6.8 Hz, 1H), 4.65 (s, 1H), 4.34 (t, *J* = 7.0 Hz, 2H), 3.98 (d, *J* = 12.3 Hz, 1H), 3.84 (d, *J* = 12.1 Hz, 1H), 3.74-3.70 (m, 1H), 3.42-3.38 (m, 1H), 1.61-1.57 (m, 2H), 1.54 (s, 3H), 1.40-1.32 (m, 4H); ^{13}C NMR (D_2O, 126 MHz) δ 136.4, 123.9 (dd, *J* = 8.0, 2.3 Hz), 98.5, 72.5, 63.2, 62.2 (d, *J* = 5.3 Hz), 29.9, 24.5, 19.0, 13.4; ^{31}P NMR (D_2O, 101 MHz) δ -5.62 (d, *J* = 20.9 Hz), -7.57 (d, *J* = 20.8 Hz).

(Z)-(16). ^{1}H NMR (D_2O, 500 MHz) δ 5.52 (t, *J* = 6.8, 1H), 4.65 (s, 1H), 4.31 (t, *J* = 7.1 Hz, 2H), 3.98 (d, *J* = 12.3 Hz, 1H), 3.84 (d, *J* = 12.1 Hz, 1H), 3.74-3.70 (m, 1H), 3.42-3.38 (m, 1H), 1.64

(s, 3H), 1.61-1.57 (m, 2H), 1.40-1.32 (m, 4H); ^{13}C NMR (D_2O, 126 MHz) δ 136.3, 125.8 (d, J = 8.6 Hz), 98.6, 72.5, 63.2, 61.8 (d, J = 5.1 Hz), 29.9, 24.5, 20.8, 19.0; ^{31}P NMR (D_2O, 101 MHz) δ –5.69 (d, J = 20.8 Hz), –7.68 (d, J = 20.8 Hz).

(E,Z)-1-Hydroxy-2-methyl-but-2-enyl diphosphate triammonium salt (8) (Davisson *et al.*, 1986).

268 mg (0.675 mmol) of protected pyrophosphate **16** are dissolved in 2.0 ml of D_2O and the pH is adjusted to 1 by addition of 40 µl of 37% DCl in D_2O. After 1 min at this pH the solution is neutralized by addition of 40 µl of 40% NaOD in D_2O and an ^1H NMR is measured that demonstrated 50% deprotection. The procedure is repeated until deprotection is finished and just small amounts of decomposition product are formed to get in total 7 min at pH 1. Purification is performed by loading the neutral solution that is diluted by addition of 2 ml of isopropanol/100 mM NH_4HCO_3 1:1 on a cellulose column (isopropanol/100 mM NH_4HCO_3 1:1, 2 × 10.5 cm) to yield 193 mg (0.616 mmol, 91%) of a white solid of **(E)-8/(Z)-8** = 7:1.

(E)-(8). ^1H NMR (D_2O, 500 MHz) δ 5.51 (tq, J = 6.8, 1.2 Hz, 1H), 4.41 (t, J = 7.2 Hz, 2H), 3.90 (s, 2H), 1.59 (s, 3H); ^{13}C NMR (D_2O, 126 MHz) δ 139.8, 120.6 (d, J = 7.7 Hz), 66.5, 62.4 (d, J = 5.3 Hz), 13.2; ^{31}P NMR (D_2O, 101 MHz) δ –4.48 (d, J = 20.8 Hz), –7.06 (d, J = 20.8 Hz).

(Z)-(8). ^1H NMR (D_2O, 500 MHz) δ 5.49 (tm, J = 6.8 Hz, 1H), 4.41 (t, J = 7.3 Hz, 2H), 4.03 (s, 2H), 1.70 (s, 3H); ^{13}C NMR (D_2O, 126 MHz) δ 139.8, 123.5 (d, J = 7.7 Hz), 61.7 (d, J = 5.1 Hz), 59.9, 20.6; ^{31}P NMR (D_2O, 101 MHz) δ –4.48 (d, J = 20.8 Hz), –7.06 (d, J = 20.8 Hz).

Reagents and conditions (steps (a) to (f) in Fig. 4: 1): (a) DHP, PPTS, 25 °C (2.5 h); (b) Ph_3PCHCO_2Et, toluene, reflux (39 h); (c) (1) DIBAH, CH_2Cl_2, –78 °C (3 h), (2) 1 M NaOH/H_2O; (d) *p*-TsCl, DMAP, CH_2Cl_2, 25 °C (1 h); (e) $((CH_3CH_2CH_2CH_2)_4N)_3HP_2O_7$, MeCN, 25 °C (2 h); (f), HCl/H_2O pH 1, 25 °C (7 min).

Example 25

Identification of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate

The structure of the GcpE product is further analyzed by comparison with the chemical shifts of a synthetic sample of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.

For this purpose, [2-^{14}C]1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate (0.36 µCi) is added to a cell extract obtained from bioengineered *Escherichia coli* cells endowed with artificial gene constructs expressing *xylB*, *ispC*, *ispD*, *ispE*, *ispF* and *gcpE* gene which are supplied with [U-$^{13}C_5$]-1-deoxy-D-xylulose (see example 16). The supernatant of the cell extract is purified by HPLC (Nucleosil 5 SB, 7.5 × 250 mm, developed with a gradient of 100 mM to 250 mM NH_4HCOO, flow rate 2 ml/min, 35 min). The product is eluted at 23 min and collected. After lyophilization the residue is dissolved in D_2O (pH 6) and subjected to ^{1}H NMR analysis (Figure 3-A).

Then, 40 µl of a solution of synthetically prepared (E,Z)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate (E/Z = 7:1) (D_2O, pH 7) are added to the NMR sample and again analyzed by ^{1}H NMR spectroscopy (Figure 3-B). On the one hand, as shown in Figure 3-B, signals accounting for (E)-1-hydroxy-2-methyl-2-butenyl are selectively increased, providing evidence that the biologically produced structure is identical with the synthetically produced one, i.e. the (E)-isomer. On the other hand, the minor (Z)-isomer raises without any correlation to signals of the biologically afforded product. Figure 3-C shows the same effects after addtion of another 40 µl of solution of the synthetically prepared (E,Z)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate.

Example 26

Incorporation of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate into the lipid soluble fraction of *Capsicum annuum* chromoplasts.

Chromoplasts are isolated by a slight modification of a method described by Camara (Camara, 1985; Camara, 1993). Pericarp of red pepper (650 g) is homogenized at 4 °C in 600 ml of 50 mM Hepes, pH 8.0, containing 1 mM DTE, 1 mM EDTA and 0.4 M sucrose (buffer A). The suspension is filtered through four layers of nylon cloth (50 µm) and centrifuged (10 min, 4,500 rpm, GSA rotor) to obtain a pellet of crude chromoplasts which is homogenized in 200 ml of buffer A. The suspension is centrifuged (10 min, 4,500 rpm, GSA rotor). The pellet is homogenized and resuspended in 3 ml of 50 mM Hepes, pH 7.6, containing 1 mM DTE. The suspension is filtered through one layer of nylon cloth (50 µm).

Reaction mixtures contain 100 mM Hepes, pH 7.6, 2 mM $MnCl_2$, 10 mM $MgCl_2$, 5 mM NaF, 2 mM $NADP^+$, 1 mM NADPH, 6 mM ATP, 20 µM FAD and 2 mg of chromoplasts. 8.8 nmol of [2-^{14}C]2*C*-methyl-D-erythritol 2,4-cyclodiphosphate, [2-^{14}C]1-hydroxy-2-methyl-2-(*E*)-butenyl diphosphate or [2-^{14}C]isopentenyl diphosphate (specific concentrations 15.8 µCi/µmol) are

added and the mixtures are incubated at 30 °C overnight. The reaction is terminated by methylene chloride extraction. The organic phase is concentrated under a stream of nitrogen. Aliquots are spotted on silica gel plates (Polygram SIL-G, UV254, Macherey-Nagel, Düren, Germany). The plates are developed with hexane : ether = 6: 1 (system I) and/or hexane : toluene = 9 : 1 (system II), respectively. The chromatograms are monitored with a phosphor imager (Storm 860, Molecular dynamics, Sunnyvale, CA, USA). The R_f-values of geranylgeraniol and the carotene fraction in system I are 0.35 and 0.9, respectively. The R_f-values of β-carotene, phytoene and phytofluene in system II are 0.65, 0,60 and 0.55, respectively.

The evaluation of the chromatogramms show that radioactivity can be efficiently diverted from 1-hydroxy-2-methyl-2-(*E*)-butenyl diphosphate into the geranylgeraniol, β-carotene, phytoene and phytofluene fractions of C. annuum chromoplasts establishing 1-hydroxy-2-methyl-2-(*E*)-butenyl diphosphate as a real intermediate of the non-mevalonate pathway downstream from 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and upstream from isopentenyl diphosphate.

Example 27

Construction of a vector carrying the *ispG* (*gcpE*) and *ispH* (*lytB*) genes of *Escherichia coli* capable for transcription and expression thereof

The *E. coli* ORF *ispH* (*lytB*) (accession no. gb AE000113) from base pair (bp) position 5618 to 6568 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GCTTGCGTCGACGAGGAGAAATTAACCATGCAGATCCTGTTGGCCACC-3', 10 pmol of the primer 5'-GCTGCTCGGCCGTTAATCGACTTCACGAATATCG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.4 µg of the vector pACYC184 (Chang and Cohen 1978, NEB) and 0.7 µg of the purified PCR product are digested with *Sal*I and *Eag*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 18 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYClytB. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYClytB is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYClytB is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prismä Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000113).

The *E. coli* ORF *ispG (gcpE)* (accession no. gb AE000338) from base pair (bp) position 372 to 1204 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-CGTACCGGATCCGAGGAGAAATTAACCATGCATAACCAGGCTCCAATTC-3', 10 pmol of the primer 5'-CCCATCGTCGACTTATTTTTCAACCTGCTGAACGTC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.0 µg of the vector pACYClytB and 0.9 µg of the purified PCR product are digested with *Bam*HI and *Sal*I in order to produce DNA fragments with overlapping ends. The restriction

mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 23 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYClytBgcpE. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYClytBgcpE is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYClytBgcpE is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000338).
The DNA sequence of the vector construct pACYClytBgcpE is shown in Annex I.

The DNA and corresponding amino acid sequence of *ispH* (*lytB*) from *Escherichia coli* is shown in Annex J.

Example 28

Construction of a vector carrying the *xylB*, *dxr*, *ispD*, *ispE*, *ispF*, *ispG* and *ispH* genes of *Escherichia coli* capable for transcription and expression of D-xylulokinase, DXP reductoisomerase, CDP-ME synthase, CDP-ME kinase cMEPP synthase, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase and IPP/DMAPP synthase

The *E. coli* ORFs ispG (formerly *gcpE*) and *ispH* (formerly *lytB*) are amplified by PCR using the plasmid pACYClytBgcpE (see example 27) as template. The reaction mixture contains 10 pmol of the primer 5'-GCGGGAGACCGCGGGAGGAGAAATTAACCATGCATAACCAGGCTCCAATTCAACG', 10 pmol of the primer 5'-AGGCTGGCGGCCGCTTAATCGACTTCACGAATATCG-3', 2 ng of pACYCgcpElytB DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 150 sec at 72 °C followed. After further incubation for 20 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

1.7 µg of the vector pBScyclo (Example 5) and 1.3 µg of the purified PCR product are digested with *Sac*II and *Not*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

22 ng of the purified vector DNA and 19 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pBScyclogcpElytB2. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pBScyclogcpElytB2 is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pBScyclogcpElytB2 is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. The DNA sequence of the vector construct pBScyclogcpElytB2 is shown in Annex K.

Example 29

Incorporation experiment with recombinant *Escherichia coli* XL1-pBScyclogcpElytB2 using [U-$^{13}C_5$]1-deoxy-D-xylulose

0.1 litre of Terrific Broth (TB) medium containing 18 mg of ampicillin are inoculated with *E. coli* strain Xl1-Blue harbouring the plasmid pBScyclogcpElytB2. The cells are grown in a shaking culture at 37 °C for overnight. At an optical density (600 nm) of 1.3-1.7 a solution containing 2.4 g of lithium lactate (25 mmol), 10 ml of crude [U-$^{13}C_5$]1-deoxy-D-xylulose (0.05 mmol) (see example 8) at a final volume of 30 ml (pH=7.4) are added continously within 2 hours. Aliquots

of 40 ml are taken at time intervals of 30 minutes and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 ml of a mixture of methanol and D_2O (6:4, v/v) containing 10 mM NaF, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the cell free extracts are recorded directly with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). In order to avoid degradation during work-up, the structures of the products are determined by NMR spectroscopy without further purification.

Example 30

Structure determination of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)

The ^{1}H-decoupled ^{13}C NMR spectrum using [U-$^{13}C_5$]1-deoxy-D-xylulose as starting material (see examples 8 and 30) displays five intense ^{1}C-^{13}C coupled signals belonging to 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate (Herz et al., 2000) and five ^{13}C-^{13}C coupled signals with low intensities belonging to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (see example 18) (100:3 ratio for the 2-methyl ^{13}C NMR signal intensities of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, respectively).

In addition a set of , five ^{13}C-^{13}C coupled signals at 21.6 (doublet), 37.8 (triplet), 64.1 (doublet), 111.6 (doublet), and 143.3 ppm (doublet of triplets) (unknown metabolite A) accompanied by signals at 21.1 (doublet), 39.6 (triplet), 59.3 (doublet), 111.8 (doublet), and 143.2 ppm (doublet of triplets) (unknown metabolite B) is detected. The ratio of the 2-methyl signal of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and the putative methyl signals of the unknown compounds at 21.6 ppm (metabolite A) and 21.1 ppm (metabolite B) is 100:24:4, respectively.

Moreover, ^{13}C coupled signals with low intensities belonging to another unknown compound (metabolite C) at 17.1 (doublet), 24.9 (doublet), 62.7 (doublet), 119.6 (double-doublet) and 139.4 ppm (multiplet) are detected. The ratio of the intensities of the putative methyl signals at 21.6 (metabolite A), 17.1 and 24.9 (metabolite C) is 100:13:13, respectively.

The ^{31}P NMR spectrum of the reaction mixture is characterized by intense signals for 2*C*-

methyl-D-erythritol 2,4-cyclodiphosphate (Herz et al., 2000). Furthermore, $^{31}P^{31}P$ coupled broadened signals are observed at a chemical shift range typical for organic diphosphates (-6 to –13 ppm, $^{31}P^{31}P$ coupling constants, 20 Hz).

Metabolite A:

The signals of metabolite A at 111.6 and 143.3 ppm are conducive of a double bond motif, and the signals at 64.1, 37.8 and 21.6 ppm reflect three aliphatic carbon atoms one of which (signal at 64.1 ppm) appears to be connected to OH or OR (R=unknown).

Additional information about the structure of the unknown metabolite A can be gleaned from the ^{13}C coupling pattern. Three of the ^{13}C NMR signals (21.6, 64.1 and 111.6 ppm) are split into doublets indicating three ^{13}C atoms each connected to only one ^{13}C-labelled neighbour, one signal (37.8 ppm) displays a pseudo-triplet signature indicating a ^{13}C atom with two adjacent ^{13}C atoms, and one signal (143.3 ppm) is split into a doublet of triplets indicating a ^{13}C atom with three ^{13}C connections. In conjunction with the chemical shifts, this connectivity pattern establish metabolite A as an isopentenyl derivative.

The complex signature for the signal at 143.3 ppm deserves a more detailed analysis. The large coupling (71 Hz) is typical for $^{13}C^{13}C$ couplings between carbon atoms involved in carbon-carbon double bonds. A 71 Hz coupling is also found for the doublet signal at 111.6 ppm representing the second carbon of the double bond. Due to the coupling pattern and the chemical shifts the presence of an exo-methylene function is obvious. The two additional ^{13}C couplings found in the triplet substructure of the signal at 143.3 ppm are both 41 Hz, and establish the respective carbon as the branching point of the structure.

HMQC experiments reveal the ^{1}H NMR chemical shifts, as well as ^{13}C-^{1}H and ^{1}-^{1}H spin systems. More specifically, the ^{13}C NMR signal at 111.6 ppm correlates to a ^{1}H NMR signal at 4.73 ppm, whereas the signal at 143.3 ppm gives no ^{13}C-^{1}H correlation. The signals at 64.1, 37.8, and 21.6 ppm give ^{13}C-^{1}H correlations to ^{1}H-signals at 4.00, 2.31, and 1.68 ppm, respectively. As shown by HMQC-TOCSY experiments, the proton signals at 2.31 and 4.00 are coupled, whereas the signals at 4.73 and 1.68 ppm are found as singlets in the HMQC-TOCSY experiment. The observed ^{1}H NMR chemical shifts in combination with the coupling patterns demonstrate that metabolite A is an isopentenyl derivative with a single bonded heteroatom (most plausibly O) at position 1.

The ^{31}C and ^{1}H chemical shifts of an authentic sample of isopentenyl diphosphate (IPP, measured in the same solvent mixture) are identical to the chemical shifts assigned to

metabolite A. Thus, metabolite A is identified as [U-$^{13}C_5$]IPP.

Metabolite B:

As noted above, the coupling and correlation pattern of metabolite B observed in the ^{13}C NMR signals, as well as in the HMQC and HMQC-TOCSY spectra, is virtually the same as for metabolite A (IPP) suggesting that the carbon connectivities of metabolite B and IPP are identical. As the most significant difference between the NMR data of metabolite B and IPP the ^{13}C NMR chemical shift of one doublet signal for metabolite B (59.3 ppm) corresponding to the C-1 signal of IPP (64.1 ppm) is upfield shifted by 4.9 ppm. This suggests that a phosphate moiety is missing at C-1 in metabolite B. Therefore, metabolite B is assigned as [U-$^{13}C_5$]isopentene-1-ol. Presumably, isopentene-1-ol is formed from IPP by the catalytic action of pyrophosphatases and phosphatases present in the experimental system.

Metabolite C:

As described above for metabolite A (IPP), the structure of metabolite C is assigned by NMR analysis. The ^{13}C coupling pattern of the signals attributed to metabolite C (three doublets, one double-doublet, one multiplet) suggests that the compound is another isopentane derivative. The chemical shifts observed for the double-doublet (119.6 ppm) and the multiplet (139.4 ppm) show that a carbon-carbon double bond connects C-2 (coupled to two ^{13}C neighbours) and C-3 (coupled to three ^{13}C neighbours) of the molecule.

The ^{1}H NMR chemical shifts of metabolite C are revealed by HMQC and HMQC-TOCSY experiments showing two singlets at 1.75 and 1.71 ppm, and a spin system comprising signals at 5.43 and 4.45 ppm. In conjunction with the chemical shifts, this correlation pattern shows that metabolite C is a dimethylallyl derivative.

The ^{13}C and ^{1}H NMR chemical shifts of an authentic sample of dimethylallyl diphosphate (DMAPP) are identical to the chemical shifts of the signals attributed to metabolite C. This leaves no doubt that metabolite C is [U-$^{13}C_5$]dimethylallyl diphosphate (DMAPP).

The NMR data of metabolite A (IPP) and metabolite C (DMAPP) are summarized in Tables 6 and 7.

Table 6. NMR data of isopentenyl diphosphate (IPP)

Position	Chemical shifts, ppm			Coupling constants, Hz				
	$^{1}H^{a}$	$^{13}C^{b}$	$^{31}P^{c}$	J_{PC}	J_{HH}	J_{PP}	J_{PH}	J_{CC}^{d}
1	4.00	64.1		4.9	6.6		6.6	34
2	2.31	37.8		8.0	6.7			40, 40
3		143.3						71, 41, 41
4	4.73	111.6						71
5	1.68	21.6						41
P			- 7.8			nd		
P			- 11.9			19,5		

[a]referenced to external trimethylsilylpropane sulfonate.

[b]referenced to external trimethylsilylpropane sulfonate.

[c]referenced to external 85 % orthophosphoric acid.

[d]observed with [U-$^{13}C_5$]IPP

Table 7. NMR data of dimethylallyl diphosphate (DMAPP)

Position	Chemical shifts, ppm			Coupling constants, Hz				
	$^{1}H^{a}$	$^{13}C^{b}$	$^{31}P^{c}$	J_{PC}	J_{HH}	J_{PP}	J_{PH}	J_{CC}^{d}
1	4.45	62.7		3.6	6.6		6.6	47
2	5.43	119.6		9.0	7.2			75, 48
3		139.4						nd
4	1.75	24.9						42
5	1.71	17.1						41
P			- 9.1			21.7		
P			- 6.4			21.5		

[a]referenced to external trimethylsilylpropane sulfonate.

[b]referenced to external trimethylsilylpropane sulfonate.

[c]referenced to external 85 % orthophosphoric acid.

[d]observed with [U-$^{13}C_5$]DMAPP

Example 31

Cloning of the *ispG* gene (fragment) from *Arabidopsis thaliana*

RNA is isolated from 1 g of 2 weeks old *Arabidopsis thaliana* var. Columbia plants (stems and leafs) by published procedures (Logemann et al. 1987).

A mixture containing 2.75 µg RNA, 50 nmol dNTP's, 1 µg random hexameric primer, 1 µg T_{15}-primer and 20 % first strand 5x buffer (Promega) in a total volume of 50 µl is incubated for 5 min. at 95 °C, cooled on ice and 500 U M-MLV reverse transcriptase (Promega) are added. The mixture is incubated for 1 h at 42 °C. After incubation at 92 °C for 5 min, RNase A (20 U) and RNase H (2 U) are added and the mixture is incubated for 30 min. at 37 °C.

The resulting cDNA (1 µl of this mixture) is used for the amplification of *ispG* by PCR.

The *A. thaliana* ORF *ispG* (accession no. dbj AB005246) without the coding region for the putative leader sequence from basepair (bp) position 2889 to 6476 is amplified by PCR using cDNA froom *A. thaliana* as template. The reaction mixture contains 25 pmol of primer CCTGCATCCGAAGGAAGCCC, 25 pmol of primer CAGTTTTCAAAGAATGGCCC, 1 µl of cDNA, 2 U of Taq DNA polymerase (Eurogentec, Seraing, Belgium) and 20 nmol of dNTPs in a total volume of 100 µl in 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 95 °C. Then 40 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 20 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis. The PCR amplificate is purified with the PCR purification kit from Qiagen. 1.7 µg of purified PCR product are obtained.

The PCR amplificate is used as template for a second PCR reaction. The reaction mixture contains 25 pmol of primer TGAATCAGGATCCAAGACGGTGAGAAGG, 25 pmol of primer TCCGTTTGGTACCCTACTCATCAGCCACGG, 2 µl of the first PCR amplification , 2 U of Taq DNA polymerase (Eurogentec, Seraing, Belgium) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 95 °C. Then 40 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C follow. After further incubation for 20 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with PCR purification kit from Qiagen. 1.4 µg of purified PCR product are obtained. 2.0 µg of the vector pQE30 and 1.4 µg of the purified PCR product are digested with *BamH*I and *Kpn*I in order to produce cohesive ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of vector DNA and 12 ng of PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pQEgcpEara. The ligation mixture is incubated over night at 4 °C. 2 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue and M15[pREP4] (Zamenhof *et. al.*, 1972) cells. The plasmid pQEgcpEara is isolated as described above. 7 µg of plasmid DNA are obtained.

The DNA insert of the plasmid pQEgcpEara is sequenced as described above. The DNA sequence is found not to be identical with the sequence in the database (accession no. dbj AB005246, see Annex L).

Example 32

Screening of IspG (GcpE) enzyme activity

0.2 g cells of XL1-pACYClytBgcpE are suspended in 1 ml 50 mM Tris hydrochloride, pH 7.4 and 2 mM DTT, cooled on ice and sonified 3 × 7 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 80 % duty cycle output, control value of 4 output. The suspension is centrifuged at 14000 rpm for 15 minutes. The supernatant is used as crude cell extract in assays described as follows.

The assay mixture contains 100 mM Tris hydrochloride, pH 7.4, 1.2 mM dithiothreitol, 10 mM NaF, 1 mM $CoCl_2$, 2 mM NADH, 20 mM (18 µCi mol^{-1}) [2-^{14}C]2*C*-methyl-erythritol 2,4-cyclodiphophate, 0.5 mM pamidronate and 100 µl crude cell extract of XL1-pACYClytBgcpE in a total volume of 150 µl. The mixture is incubated for 10 to 45 min at 37 °C and cooled on ice. 10 µl of 30 % (g/v) trichloroacetic acid are added and the mixture is neutralized with 20 µl of 1 M NaOH. The mixture is centrifuged at 14.000 rpm for 10 minutes. Aliquotes of 130 µl of the supernatant are analyzed by reversed-phase ion-pair HPLC using a column of Multospher 120 RP 18-AQ-5 (4.6 × 250 mm, CS-Chromatographie Service GmbH, Langerwehe, Germany).

The column is developed with a linear gradient of 7-21 % (v/v) methanol in 20 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0 at a flow rate of 1 ml min^{-1} and further with a linear gradient of 21-49 % (v/v) methanol in 15 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0. After washing the column with 49 % (v/v) methanol in 5 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0, the column is equilibrated with 7 % (v/v) methanol in 20 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0. The effluent is monitored by a continous-flow radio detector (Beta-RAM, Biostep GmbH, Jahnsdorf, Germany). The retention volumes of 2*C*-methyl-erythritol 2,4-cyclodiphophate, 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate, DMAPP/ IPP are 18, 24 and 39 ml respectively.

After 10 minutes of incubation, about 13 % of 2*C*-methyl-erythritol 2,4-cyclodiphophate have been converted into 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate (5 %) and into DMAPP/IPP (8 %), respectively.
After 45 min, no 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate, but about 21 % of DMAPP/IPP was found in the assay mixture.

Example 33

Screening of IspH (LytB) activity

Assay mixtures contain 100 mM Tris hydrochloride, pH 7.4, 1.2 mM DTT, 10 mM NaF, 0.5 mM NADH, 60 µM FAD, 0.004 µM (18 µCi $µmol^{-1}$) [2-^{14}C]1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate, 0.5 mM pamidronate (Dunford *et al.*, 2001) and 20 µl of crude cell extract of M15-pMALlytB cells (prepared as described in example 2) in a total volume of 150 µl. The mixture is incubated for 30 min at 37 °C. The reaction is terminated by cooling on ice, addition of 10 µl of 30 % (g/v) trichloroacetic acid and immediate neutralization with 20 µl 1 M sodium hydroxide. The mixtures are centrifuged and aliquots (130 µl) of the supernatant are analyzed by reversed-phase ion-pair HPLC using a column of Multospher 120 RP 18-AQ-5 (4.6 x 250 mm, CS-Chromatographie Service GmbH, Langerwehe, Germany) analyzed by reversed-phase ion-pair HPLC using a column of Multospher 120 RP 18-AQ-5 (4.6 × 250 mm, CS-Chromatographie Service GmbH, Langerwehe, Germany). The column is developed with a linear gradient of 7-21 % (v/v) methanol in 20 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0 at a flow rate of 1 ml min^{-1} and further with a linear gradient of 21-49 % (v/v) methanol in 15 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0. After

washing the column with 49 % (v/v) methanol in 5 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0, the column is equilibrated with 7 % (v/v) methanol in 20 ml of 10 mM tetra-n-butylammonium hydrogen phosphate, pH 6.0. The effluent is monitored by a continous-flow radio detector (Beta-RAM, Biostep GmbH, Jahnsdorf, Germany).

Under standard assay conditions, the HPLC peak corresponding to the substrate 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate is completely diminished, whereas two new peaks corresponding to DMAPP and IPP appear, when crude cell extract of *E. coli* M15-pMALlytB cells is used as protein source. No conversion of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate into DMAPP and IPP can be observed, when crude cell extract of *E. coli* wild-type is used as protein source. This findings clearly show that the FAD and NADH- or NADPH-dependent conversion of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate into DMAPP and IPP is catalyzed by the recombinant LytB protein. The addition of pamidronate in the assay mixtures prevents a further metabolization of IPP and DMAPP by highly active prenyl transferases present in crude *E. coli* extracts and affects therefore the complete conversion of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate into DMAPP and IPP.

Example 34

Construction of a vector carrying the *dxs, xylB* and *ispC* genes capable for the transcription and expression thereof

The *B. subitis* ORF *dxs* (accession no. dbj D84432) from base pair (bp) position 193991 to 195892 is amplified by PCR using pBSDXSBACSU plasmid DNA as template (see patent application PCT/EP00/07548). The reaction mixture contains 10 pmol of the primer 5'-GGCGACTCGCGAGAGGAGAAATTAACCATGGATCTTTTATCAATACAGGACC-3', 10 pmol of the primer 5'-GGCACCCGGCCGTCATGATCCAATTCCTTTGTGTG-3', 20 ng DNA of pBSDXSBACSU plasmid, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 120 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the

mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis. The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.4 µg of the vector pACYC184 (Chang and Cohen 1978, NEB) and 1.8 µg of the purified PCR product are digested with *Nru*I and *Eag*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 19 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYCdxs. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYCdxs is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYCdxs is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (dbj D84432).

2.0 µg of the vector pACYCdxs and 8 µg of the vector pBScyclo (see example XXx) are digested with *Eag*I and *Sal*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the digested and purified pACYCdxs vector DNA and 30 ng of a by DNA electrophoresis separated and purified 2.7 kb *Eag*I/*Sal*I fragment (containing the ORFs *xylB* and *ispC* from *E. coli*) are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYCdxsxylBispC. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYCdxsxylBispC is isolated with the

plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYCdxsxylBispC is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions.

Example 35

Construction of a vectors carrying the *dxs, xylB*, *ispC*, and *ispG* and optionally *ispH* genes capable for the transcription and expression thereof

The *E. coli* ORF *ispH* (*lytB*) (accession no. gb AE000113) from base pair (bp) position 5618 to 6568 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GCTTGCGTCGACGAGGAGAAATTAACCATGCAGATCCTGTTGGCCACC-3', 10 pmol of the primer 5'-GCTGCTCTCGAGTTAATCGACTTCACGAATATCG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.5 µg of the vector pACYCdxsxylBispC (see example 34) are linearized with *Sal*I and 0.9 µg of the purified PCR product are digested with *Sal*I and *Xho*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

15 ng of the purified vector DNA and 18 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pACYCdxsxylBispClytB. The ligation mixture is incubated for 2 h at 25 °C.

1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pACYCdxsxylBispClytB is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pACYCdxsxylBispClytB is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000113).

The *E. coli* ORF *ispG (gcpE)* (accession no. gb AE000338) from base pair (bp) position 372 to 1204 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GGTCGAGTCGACGAGGAGAAATTAACCATGCATAACCAGGCTCCAATTC-3', 10 pmol of the primer 5'-CCCATCCTCGAGTTATTTTTCAACCTGCTGAACGTC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.
The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

Each 2.0 µg of the vectors pACYCdxsxylBispC (see example 34) and pACYCdxsxylBispClytB (see above) are linearized with *Sal*I and 1.1 µg of the purified PCR product are digested with *Sal*I and *Xho*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

18 ng of the purified vector DNAs and 23 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmids pACYCdxsxylBispCgcpE and pACYCdxsxylBispClytBgcpE. The ligation mixtures are incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into

electrocompetent *E. coli* XL1-Blue cells. The plasmids pACYCdxsxylBispCgcpE and pACYCdxsxylBispClytBgcpE are isolated with the plasmid isolation kit from Qiagen.

The DNA inserts of the plasmids pACYCdxsxylBispCgcpE and pACYCdxsxylBispClytBgcpE are sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. They are identical with the DNA sequence of the database entry (gb AE000338).

Example 36

Incorporation experiment with recombinant *Escherichia coli* XL1-pACYCdxsxylBispCgcpE using [U-$^{13}C_6$]glucose

0.2 litre of Terrific Broth (TB) medium containing 5 mg of chloramphenicol are inoculated with *E. coli* strain XL1-Blue harbouring the plasmid pACYCdxsxylBispCgcpE. The cells are grown in a shaking culture at 37 °C overnight. At an optical density (600 nm) of 1.7-2.4 a solution containing 1 g of lithium lactate (10 mmol), 200 mg [U-$^{13}C_6$]glucose (1.1 mmol) at a final volume of 24 ml (pH=7.4) are added continously within 2 hours. Then, after 1 hour an aliquot of 40 ml was taken and centrifuged for 20 min at 5,000 rpm and 4 °C. The cells are washed with water containing 0.9 % NaCl and centrifuged as described above. The cells are suspended in 700 µl of a mixture of methanol-d_4 and D_2O (6:4, v/v) containing 10 mM NaF, cooled on ice and sonified 3 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 90 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15,000 rpm for 15 min. NMR spectra of the cell free extracts are recorded directly with a Bruker AVANCE DRX 500 spectrometer (Karlsruhe, Germany). In order to avoid degradation during work-up, the structures of the products are determined by NMR spectroscopy without further purification.

The ^{13}C-NMR spectra showed signals accounting for 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate (cf. Tables 2 and 4, example 18) as major product. A formation of 2*C*-methyl-D-erythritol 2,4-cylodiphosphate could not be detected.

Example 37

Incorporation experiment with recombinant *Escherichia coli* XL1-pACYCdxsxylBispClytBgcpE using glucose

Example 36 can be carried out with recombinant *Escherichia coli* XL1-pACYCdxsxylBispClytBgcpE using glucose for converting glucose to isopentenyl diphosphate and/or dimethylallyl diphosphate.

Example 38

Cloning of the *ispG* gene of *Escherichia coli* and expression as maltose binding fusion protein (MBP-IspG)

The *E. coli* ORF *ispG (gcpE)* (accession no. gb AE000338) from base pair (bp) position 372 to 1204 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-GAACCGGAATTCATGCATAACCAGGCTCCAATTC-3', 10 pmol of the primer 5'-CGAGGCGGATCCCATCACG-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 60 sec at 94 °C, 60 sec at 50 °C and 90 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.2 µg of the vector pMAL-C2 (NEB) and 0.8 µg of the purified PCR product are digested with *EcoR*I and *Bam*HI in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR purification kit from Qiagen.

20 ng of the purified vector DNA and 15 ng of the purified PCR product are ligated together

with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pMALgcpE. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pMALgcpE is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pMALgcpE is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000338).

Example 39

Cloning of the *ispH* gene of *Escherichia coli* and expression as maltose binding fusion protein (MBP-IspH)

The *E. coli* ORF *ispH* (*lytB*) (accession no. gb AE000113) from base pair (bp) position 5618 to 6568 is amplified by PCR using chromosomal *E. coli* DNA as template. The reaction mixture contains 10 pmol of the primer 5'-TGGAGGGGATCCATGCAGATCCTGTTGGCCACC-3', 10 pmol of the primer 5'-GCATTTCTGCAGAACTTAGGC-3', 20 ng of chromosomal DNA, 2 U of Taq DNA polymerase (Eurogentec) and 20 nmol of dNTPs in a total volume of 100 µl containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-hydrochloride, pH 8.8 and 0.1 % (w/w) Triton X-100.

The mixture is denaturated for 3 min at 94 °C. Then 30 PCR cycles for 45 sec at 94 °C, 45 sec at 50 °C and 60 sec at 72 °C followed. After further incubation for 10 min at 72 °C, the mixture is cooled to 4 °C. An aliquot of 2 µl is subjected to agarose gel electrophoresis.

The PCR amplificate is purified with the PCR purification kit from Qiagen (Hilden).

2.2 µg of the vector pMAL-C2 (NEB) and 0.7 µg of the purified PCR product are digested with *BamH*I and *Pst*I in order to produce DNA fragments with overlapping ends. The restriction mixtures are prepared according to the conditions supplied by the customer (NEB) and are incubated for 3 h at 37 °C. Digested vector DNA and PCR product are purified using the PCR

purification kit from Qiagen.

20 ng of the purified vector DNA and 14 ng of the purified PCR product are ligated together with 1 U of T4-Ligase (Gibco), 2 µl of T4-Ligase buffer (Gibco) in a total volume of 10 µl, yielding the plasmid pMALlytB. The ligation mixture is incubated for 2 h at 25 °C. 1 µl of the ligation mixture is transformed into electrocompetent *E. coli* XL1-Blue cells. The plasmid pMallytB is isolated with the plasmid isolation kit from Qiagen.

The DNA insert of the plasmid pMALlytB is sequenced by the automated dideoxynucleotide method using an ABI Prism 377™ DNA sequencer from Perkin Elmer with the ABI Prism™ Sequencing Analysis Software from Applied Biosystems Divisions. It is identical with the DNA sequence of the database entry (gb AE000113).

Example 40

Preparation and purification of recombinant IspG maltose binding fusion protein (MRP-IspG)

0.5 liter of Luria Bertani (LB) medium containing 90 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harboring plasmid pMALgcpE. The culture is grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.7, the culture is induced with 2 mM IPTG. The culture is grown for further 5 h. The cells are harvested by centrifugation for 20 min at 5,000 rpm and 4 °C. The cells are washed with 20 mM Tris hydrochloride pH 7.4, centrifuged as above and frozen at -20 °C for storage.

2 g of the cells are thawed in 20 ml of 20 mM Tris hydrochloride pH 7.4, 0.2 M sodium chloride and 0.02 % (g/v) sodium acide (buffer A) in the presence of 1 mg ml^{-1} lysozyme and 100 µg ml^{-1} DNaseI. The mixture is incubated at 37 °C for 30 min, cooled on ice and sonified 6 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 70 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15, 000 rpm at 4 °C for 30 min. The cell free extract is applied on a column of amylose resin FF (column volume 25 ml, NEB) previously equilibrated with buffer A at a flowrate of 2 ml min^{-1}. The column is washed with 130 ml of buffer A. MRP-IspG is eluted with a linear gradient of 0-10 mM maltose in buffer A. MRP-IspG containing fractions are combined according to SDS-PAGE and dialyzed overnight against 100 mM Tris hydrochloride pH 7.4. The homogeneity of MRP-IspG is judged by SDS-

PAGE. One band at 84 kDa is visible, which is in line with the calculated molecular mass. The yield of pure MRP-IspG is 9 mg.

Example 41

Preparation and purification of recombinant IspH maltose binding fusion protein (MRP-IspH)

0.5 liter of Luria Bertani (LB) medium containing 90 mg of ampicillin are inoculated with 10 ml of an overnight culture of *E. coli* strain XL1-Blue harboring plasmid pMALlytB. The culture is grown in a shaking culture at 37 °C. At an optical density (600 nm) of 0.7, the culture is induced with 2 mM IPTG. The culture is grown for further 5 h. The cells are harvested by centrifugation for 20 min at 5,000 rpm and 4 °C. The cells are washed with 20 mM Tris hydrochloride pH 7.4, centrifuged as above and frozen at -20 °C for storage.

2 g of the cells are thawed in 20 ml of 20 mM Tris hydrochloride pH 7.4, 0.2 M sodium chloride and 0.02 % (g/v) sodium acide (buffer A) in the presence of 1 mg ml^{-1} lysozyme and 100 µg ml^{-1} DNaseI. The mixture is incubated at 37 °C for 30 min, cooled on ice and sonified 6 x 10 sec with a Branson Sonifier 250 (Branson SONIC Power Company) set to 70 % duty cycle output, control value of 4 output. The suspension is centrifuged at 15, 000 rpm at 4 °C for 30 min. The cell free extract is applied on a column of amylose resin FF (column volume 25 ml, NEB) previously equilibrated with buffer A at a flowrate of 2 ml min^{-1}. The column is washed with 130 ml of buffer A. MRP-IspH is eluted with a linear gradient of 0-10 mM maltose in buffer A. MRP-IspH containing fractions are combined according to SDS-PAGE and dialyzed overnight against 100 mM Tris hydrochloride pH 7.4. The homogeneity of MRP-IspH is judged by SDS-PAGE. One band at 78 kDa is visible, which is in line with the calculated molecular mass. The yield of pure MRP-IspH is 14 mg.

Example 42

Synthesis of 1-hydroxy-2-methyl-but-2-enyl-4-diphosphate (see Fig. 7)

4-Chloro-2-methyl-2-buten-1-al (Choi et al. (1999) J. Org. Chem. 64, 8051-8053)
A solution containing 1.17 ml of 2-methyl-2-vinyl-oxirane (12 mmol), 1.6 g of $CuCl_2$ (12 mmol) and 510 mg of LiCl (12 mmol) in 10 ml of ethylactetate was heated to 80 °C for 30 min. The

reaction was stopped by adding 50 g of ice. The mixture was filtered through a sintered glass funnel under reduced pressure. 100 ml of CH_2Cl_2 was added and the organic phase was separated. The aqueous layer was extracted two times with 100 ml of CH_2Cl_2. The combined organic phase was dried over anhydrous $MgSO_4$, filtered, and concentrated. The crude product was purified by chromatography over silica gel (CH_2Cl_2, 3 x 37 cm) to yield 0.755 g of a yellow liquid (6.4 mmol, 53 %).

^{1}H NMR ($CDCl_3$, 500 MHz) δ 9.43 (s, 1H), 6.50 (t, J= 7.5 Hz, 1H), 4.24 (d, J=7.5 Hz, 2H), 1.77 (s, 3H)

^{13}C NMR ($CDCl_3$, 125 MHz) δ 194.3, 145.7, 141.1, 38.6, 9.1

4-Chloro-2-methyl-2-buten-1-al-dimethyl-acetal

A solution of 184 mg 4-chloro-2-methyl-2-buten-1-al (1.55 mmol), 600 µl of $HC(OMe)_3$ (5.6 mmol) and a catalytic amount of p-TsOH was incubated for 3 h at roomtemperature. The crude mixture was purified by chromatography over silica gel (n-hexane/ethylacetate 7:3) to yield 177 mg of a colourless liquid (1.08 mmol, 72 %).

^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.78 (t, 1H, J=7.9), 4.47 (s, 1H), 4.15 (d, J=7.9 Hz, 2H), 3.33 (s, 6 H), 1.73 (s, 3H)

^{13}C NMR ($CDCl_3$ 125 MHz) δ 137.6, 124.4, 106.0, 53.5, 39.6, 11.4

(*E*)-3-Formyl-2-buten-1-diphosphate triammonium salt (Davisson et al. (1986) *J. Org. Chem.*, *51*, 4768)

To a solution of 4-chloro-2-methyl-2-buten-1-al-dimethyl-acetal chloride (25 mg, 0.15 mmol) in 250 µl of MeCN a solution of 0.162 g (0.18 mmol) of tris(tetra-*n*-butylammonium) hydrogen pyrophosphate in 400 µL of MeCN was added slowly at room temperature, leading to an orange-red solution. After 2 h the reaction was finished and the solvent was removed under reduced pressure. The orange oil was dissolved in 3 mL of H_2O and passed through a column of DOWEX 50 WX8 (1 × 4 cm) cation-exchange resin (NH_4^+ form) that has been equilibrated with 20 mL of 25 mM NH_4HCO_3. The column was eluted with 20 mL of 25 mM NH_4HCO_3. The resulting solution was lyophilized. The obtained solid was dissolved in 2 ml water and acidified with aqueouus HCl to pH=3. After 2 minutes the solution was neutralized and lyophylisized.

^{1}H NMR (D_2O, 360 MHz) δ 9.37 (s, 1H), 6.86 (t, 1H, 5.6Hz), 4.85 (dd, J=7.9, J=5.8 Hz, 2H), 1.72 (s, 3H)

^{13}C NMR (D_2O, 90 MHz) δ 199.2, 153.1 (d, J=7.5 Hz), 138.5, 63.2 (d, J=4.9), 8.5

[1-^{3}H]1-hydroxy-2-methyl-but-2-enyl-4-diphosphate

A solution containing 50 mCi (15 µmol) $NaBH_3T$, 15 µmol 3-formyl-2-buten-1-diphosphate triammonium salt and 100 mM Tris/HCl pH=8 was incubated for 30 minutes at room temperature. The solution was acidified by adding 1 M HCl to pH=2. After 2 minutes the solution was neutralizied by adding 1 M NaOH.

The product was characterizied by ion-exchange chromatography (see examples 20 and 25).

Example 43

γδ T cell stimulation assays

PBMCs from healthy donors (donor A and donor B) are isolated from heparinized peripheral blood by density centrifugation over Ficoll-Hypaque (Amersham Pharmacia Biotech, Freiburg, Germany). 5 × 10^5 PBMCs/well are cultivated in 1 mL RPMI 1640 medium supplemented with 10% human AB serum (Klinik rechts der Isar, München, Germany), 2 mM L-glutamine, 10 µM mercaptoethanol. Amounts of recombinant human IL-2 (kindly provided by Eurocetus, Amsterdam, The Netherlands) and substrates are varied from 1 to 10 U and 10 to 0.1 µM, respectively. 20 µM IPP (Echelon, Research Laboratories Inc., Salt Lake City, USA) serves as a positive control whereas medium alone serves as negative control. Incubation is done for seven days at 37 °C in the presence of 7% CO_2. The harvested cells are double-stained with fluorescein isothiocyanate (FITC)-conjugated mouse anti-human monoclonal antibody Vδ2 TCR and phycoerythrin (PE)-conjugated monoclonal CD3 antibody. The cells are analyzed using a FACScan supported with Cellquest (Becton Dickinson, Heidelberg, Germany).

The substrates (E)-1-hydroxy-3-methyl-but-2-enyl 4-diphosphate (HMBPP) and 3-formyl-but-2-enyl 1-diphosphate (Aldehyd) were prepared synthetically as described above.

It is found that both synthetically prepared substrates (HMBPP and Aldehyd) show at least double stimulation compared to IPP when used at a concentration that is 200-fold lower than the concentration of the IPP sample (Table 8).

Table 8. Activation of γδT-cells by phosphororganic compounds

Substrate	Concentration [μM]	IL-2 [U]	% γδT-cells Donor A	Donor B
Medium	-	1	1.51	2.75
Medium	-	5	1.45	2.23
Medium	-	10	1.32	1.68
IPP	20	1	8.19	6.24
IPP	20	5	14.42	9.32
IPP	20	10	16.6	11.86
IPP	1	1	1.56	2.22
IPP	1	5	1.59	2.67
IPP	1	10	1.71	2.19
IPP	0.1	1	1.3	2.15
IPP	0.1	5	1.3	2.26
IPP	0.1	10	1.01	2.54
HMBPP	10	1	3.3	31.42
HMBPP	10	5	17.38	63.48
HMBPP	10	10	24.94	63.34
HMBPP	1	1	5.57	35.34
HMBPP	1	5	14.4	54.12
HMBPP	1	10	19.85	55.90
HMBPP	0.1	1	11.78	32.21
HMBPP	0.1	5	22.92	44.69
HMBPP	0.1	10	34.69	36.33
HMBPP/IPP	0.5/0.5	1	7	30.35
HMBPP/IPP	0.5/0.5	5	15.38	53.76
HMBPP/IPP	0.5/0.5	10	24.19	46.58
Aldehyd	10	1	12.19	30.69
Aldehyd	10	5	34.69	30.33
Aldehyd	10	10	38.99	38.85
Aldehyd	1	1	15.91	21.18
Aldehyd	1	5	40.13	30.76
Aldehyd	1	10	48.28	36.69
Aldehyd	0.1	1	10	13.54
Aldehyd	0.1	5	19.77	18.45
Aldehyd	0.1	10	21.93	25.82
Aldehyd/IPP	0.5/0.5	1	13.98	22.11
Aldehyd/IPP	0.5/0.5	5	33.94	32.06
Aldehyd/IPP	0.5/0.5	10	42.84	36.25

IPP: isopentenyl diphosphate
HMBPP: (*E*)-1-hydroxy-3-methyl-but-2-enyl 4-diphosphate
Aldehyd: (*E*)-3-formyl-but-2-enyl 1-diphosphate (prepared according to example 42)

Example 44

High through-put screening assay of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate

synthase (IspG) activity

Assay mixtures contain 20 mM potassium phosphate, pH 7.0, 0.4 mM NADH, 0.5 mM $CoCl_2$, 0.2 mM 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate, and 50 µl protein in a total volume of 1 ml. The mixtures are incubated at 37 °C. The oxidation of NADH is monitored photometrically at 340 nm. Alternatively, the concentration of NADH is determined by measuring the relative fluorescence of NADH at 340 nm excitation/460 nm emission.

Example 45

High through-put screening assay of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase (IspH) activity

Assay mixtures contain 20 mM potassium phosphate, pH 7.0, pH 8.0, 0.4 mM NADH, 20 µM FAD, 0.5 mM $CoCl_2$, 0.2 mM 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate, and 50 µl protein in a total volume of 1 ml. The mixtures are incubated at 37 °C. The oxidation of NADH is monitored photometrically at 340 nm. Alternatively, the concentration of NADH is determined by measuring the relative fluorescence of NADH at 340 nm excitation/460 nm emission.

References

Altincicek, B., Kollas, A. K., Sanderbrand, S., Wiesner, J., Hintz, M., Beck, E. & Jomaa, H. (2001). GcpE Is Involved in the 2-*C*-Methyl-D-Erythritol 4-Phosphate Pathway of Isoprenoid Biosynthesis in *Escherichia coli*. *J. Bacteriol.* **183**, 2411-2416.

Altincicek, B. *et al.* and Jomaa, H. (2001) *J. Immunology*, **166**, 3655-3658.

Arigoni D. & Schwarz M. K. (1999) Ginkgolide biosynthesis. In *Comprehensive natural product chemistry* (Barton D. and Nakanishi K., eds.), Vol. 2, pp. 367-399, Pergamon.

Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., VanLoon, A. P. G. M., Taylor, S., Campobasso, N., Chiu, H.-J., Kinsland, C., Reddick, J. J. & Xi, J. (1999) Thiamin biosynthesis in prokaryotes. *Arch. Microbiol.* **171**, 293-300.

Blagg, B. S. J. & Poulter, C. D. (1999) Synthesis of 1-deoxy-D-xylulose and 1-deoxy-D-xylulose 5-phosphate. J. Org. Chem. 64, 1508-1511.

Bullock, W. O., Fernandez, J. M., & Short, J. M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA* Escherichia coli with ß-galactosidase selection. BioTechniques **5,** 376-379.

Camara, B. *Methods in Enzymology* **1985**, eds. Law, J. H. & Rilling, H. C. (Academic Press, London) Vol. 110, pp. 267-273.

Camara, B. *Methods in Enzymology* **1993**, ed. Packer, L. (Academic Press, London) Vol. 214, pp. 352-365.

Campos, N., Rodriguez-Concepcion, M., Seemann, M., Rohmer, M. and Boronat, A. (2001) Identification of *gcpE* as a novel gene of the 2-*C*-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli. FEBS Lett.* **488**, 170-173.

Chang, A. C. Y. & Cohen, S. N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141-1145.

Cunningham, F. X. Jr., Lafond, T. P. & Gantt E. (2000). Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. *J. Bacteriol.* **182**, 5841-5848.

Davisson, V. J.; Woodside, A. B.; Neal, T. R.; Stremler, K. E.; Muehlbacher, M.; Poulter, C. D. *J. Org. Chem.* **1986**, *51*, 4768.

Dower, W. J., Miller, J. F., & Ragsdale, C. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16,** 6127-6145.

Eisenreich, W., Rohdich, F. & Bacher, A. (2001) Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Science* **6**, 78-84.

Fournié, J.J. and Bonneville, M. (1996) Res. Immunol. 147, 338-347.

Giner, J.-L. (1998) New and efficient synthetic routes to 1-deoxy-D-xylulose. *Tetrahedron Lett.* **39**, 2479-2482.

Hagiwara, H.; Uda, H. *J. Chem. Soc. Trans. I* **1984**, 91.

Hwang, C. K.; Li, W. S.; Nicolaou, K. C. *Tetrahedron Lett.* **1984**, *25*, 2295.

Herz, S., Wungsintaweekul, J., Schuhr, C. A., Hecht, S., Lüttgen, H., Sagner, S., Fellermeier, M., Eisenreich, W., Zenk, M. H., Bacher, A. & Rohdich, F. (2000). Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate to 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate. *Proc. Natl. Acad. Sci. USA* **97**, 2486-2490.

Jomaa, H., Feurle, J., Lühs, K., Kunzmann, V., Tony, H.P., Herderich, M. and Wilhelm, M.(1999) *FEMS Immunology and Medical Microbiology* 25, 371-378.

Kennedy, I. A., Hemscheidt, T., Britten, J. F., Spenser, I. D. (1995) 1-Deoxy-D-xylulose *Can. J. Chem.* **73**, 1329-1333.

Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C. A., Fellermeier, M., Sagner, S., Zenk, M. H., Bacher, A. & Eisenreich, W. (2000). Biosynthesis of terpenoids: YchB protein of *Escherichia Coli* phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2*C*-methyl-D-erythritol. *Proc. Natl. Acad. Sci. USA* **97**, 1062-1067.

Meade, H. M., Long, S. R., Ruvkun, C. B., Brown, S. E., & Auswald, F. M. (I982). Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenis. *J. Bacteriol.* **149,** 114-122.

Piel, J. & Boland, W. (1997) Highly efficient and versatile synthesis of isotopically labeled 1-deoxy-D-xylulose *Tetrahedron Lett.* **38**, 6387-6390.

Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., Eisenreich, W., Bacher, A & Zenk, M. H. (1999). Cytidine 5'-triphosphate biosynthesis of isoprenoids: YgbP protein of *Escherichia coli* catalyzes the formation of 4-diphosphocytidyl-2*C*-methylerythritol. *Proc. Natl. Acad. Sci. USA* **96**, 11758-11763.

Sanger, F., Nicklen, S., & Coulson, A. R. (1992). DNA sequencing with chain-terminating inhibitors. *Biotechnology* **24**, 104-8.

Shono, T., Matsamura, Y., Hamaguchi, H. & Naitoh, S. (1983) Synthesis of 2-methyl-3-hydroxy-4H-pyrane-4-one and 4-hydroxy-5-methyl-2H-furane-3-one from carbohydrates. *J. Org. Chem.* **48**, 5126-5128.

Takahashi, S., Kuzuyama, T., Watanabe H. & Seto, H. (1998) A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalysing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proc. Natl. Acad. Sci. USA* **95**, 9879-9884.

Watanabe, H.; Hatakeyama, S.; Tazumi, K.; Takano, S.; Masuda, S.; Okano, T.; Kobayashi, T.; Kubodera, N. *Chem. Pharm. Bull.* **1996**, *44*, 2280.Wungsintaweekul, J., Herz, S., Hecht, S., Eisenreich, W., Feicht, R., Rohdich, F., Bacher, A. & Zenk, M. H. (2001). Phosphorylation of 1-deoxy-D-xylulose by D-xylulokinase of *Escherichia coli. Eur. J. Biochem.* **268**, 310-316.

Claims

1. A protein in a form that is functional for the enzymatic conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate notably in its (*E*)-form.

2. The protein according to claim 1, wherein it is functional for said conversion in the presence of NADH and/or NADPH.

3. The protein according to claim 2, wherein it is functional for said conversion in the presence of Co^{2+}.

4. The protein according to one of claims 1 to 3, wherein it has a sequence encoded by the *ispG* (formerly *gcpE*) gene of *E. coli* or a function-conservative homologue of said sequence.

5. A protein in a form that is functional for the enzymatic conversion of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably in its (*E*)-form, to isopentenyl diphosphate and/or dimethylallyl diphosphate.

6. The protein according to claim 5, wherein it is in a form functional for said conversion in the presence of FAD and NAD(P)H.

7. The protein according to claim 6, wherein it is in a form functional for said conversion in the presence of a metal ion selected from the group of manganese, iron, cobalt, or nickel ion.

8. The protein according to one of claims 5 to 7, wherein it has a sequence encoded by the *ispH* (formerly *lytB*) gene of *E. coli* or a function-conservative homologue of said sequence.

9. The protein according to one of claims 1 to 8, wherein it is a plant protein, notably from *Arabidopsis thaliana*.

10. The protein according to one of claims 1 to 8, wherein it is a bacterial protein, notably from *E. coli*.

11. The protein according to one of claims 1 to 8, wherein it is a protozoal protein, notably from *Plasmodium falciparum*.

12. Purified isolated nucleic acid encoding the protein according to one of claims 1 to 4 and/or the protein according to one of claim 5 to 8 with or without introns.

13. A DNA expression vector containing the sequence of the nucleic acid according to claim 12.

14. Use of a protein according to one of claims 1 to 11 for screening a chemical library for an inhibitor of the biosynthesis of isoprenoids.

15. Cells, cell cultures, organisms or parts thereof recombinantly endowed with the sequence of the nucleic acid according to claim 12 or with the vector according to claim 13, wherein said cell is selected from the group consisting of bacterial, protozoal, fungal, plant, insect and mammalian cells.

16. Cells, cell cultures, organisms or parts thereof according to claim 15, wherein it is recombinantly endowed with a vector containing a nucleic acid sequence encoding a protein according to one of claims 1 to 4 and/or a protein according to one of claims 5 to 8, and wherein said cell is optionally further endowed with at least one gene selected from the following group: *dxs*, *dxr*, *ispD* (formerly *ygbP*); *ispE* (formerly *ychB*); *ispF* (formerly *ygbB*) of *E. coli* or a function-conservative homologue thereof, or a function-conservative fusion, deletion or insertion variant of any of the above genes.

17. Cells, cell cultures, or organisms or parts thereof transformed or transfected for an increased rate of formation of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably in its (*E*)-form, compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection.

18. Cells, cell cultures, or organisms or parts thereof transformed or transfected for an increased rate of conversion of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate to isopentenyl diphosphate and/or dimethylallyl diphosphate compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection.

19. Cells, cell cultures, or organisms or parts thereof according to claim 15 transformed or transfected for an increased expression level of the protein of one of claims 1 to 4 and/or the protein of one of claims 5 to 8 compared to cells, cell cultures, or organisms or parts thereof absent said transformation or transfection.

20. Cells, cell cultures or organisms or parts thereof in accordance with claim 15 or 16, characterized by the recombinant endowment with sets of genes as follows:
ispC (formerly *dxr*), *ispD*, *ispE*, *ispF*, *ispG* (formerly *gcpE*); or
ispC, *ispD*, *ispE*, *ispF*, *ispG*, *ispH* (formerly *lytB*); or
dxs, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*; or
dxs, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*; or
dxs, *ispC*, *ispG*, or
dxs, *ispC*, *ispG*, *ispH*
of *E. coli* or a function-conservative homologue thereof and/or a function-conservative fusion, deletion or insertion variant of any of the above genes.

21. Cells, cell cultures or organisms or parts thereof in accordance with claim 20, characterized by further recombinant endowment(s) with gene(s) being functional for biosynthetic steps downstream from the C5 isoprenoids.

22. Cells, cell cultures or organisms or parts thereof in accordance with one of claims 15 to 21, wherein at least one gene of said recombinant endowments is equipped with artificial ribosomal binding site(s) for expression of the corresponding gene product(s) at a rate enhanced compared to the rate in the absence of the artificial ribosomal binding site(s).

23. Cells, cell cultures or organisms or parts thereof in accordance with one of claims 15 to 22, wherein at least one of said recombinant endowments is due to a high copy replication vector.

24. Cells, cell cultures or organisms or parts thereof in accordance with one of claims 15 to 23, wherein they are of bacterial, protozoal, fungal, plant or animal origin.

25. Use of the cells, cell cultures or organisms, or parts thereof in accordance with one of claims 15 to 24 or disruption products thereof for the enhanced rate of *in vivo* formation or for the efficient *in vitro* production of an, optionally isotopically labelled, biosynthetic intermediate or product of the non-mevalonate isoprenoid biosynthetic pathway, optionally by feeding 1-deoxy-D-xylulose or glucose that may be isotopically labelled.

26. Use according to claim 25, wherein said intermediate or product is a C5-isoprenoid intermediate compound; or a >C5-isoprenoid compound; or a terpenoid compound.

27. Use according to one of claims 25 or 26, wherein the rate of formation or production is enhanced by providing a source for CTP.

28. Use according to claim 27, wherein the source for CTP is cytidine and/or uridine and/or cytosine and/or uracil and/or ribose and and/or ribose 5-phosphate and/or any biosynthetic precursor of CTP.

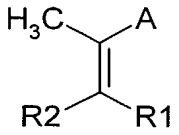
29. Use according to one of claims 25 to 28, wherein the rate of formation or production is enhanced by providing a source for phosphorylation enhancement.

30. Use according to claim 29, wherein the source for phosphorylation enhancement is glycerol 3-phosphate and/or phosphoenolpyruvate and/or inorganic phosphate and/or inorganic pyrophosphate and/or any organic phosphate or pyrophosphate.

31. Use according to one of claims 25 to 30, wherein the rate of formation or production is enhanced by providing a source for reduction equivalents.

32. Use according to claim 31, wherein the source for reduction equivalents is succinate and/or lipids and/or glucose and/or glycerol and/or lactate.

33. Optionally isotope-labelled compound of the following formula I or a salt thereof:



(I)

whereby R^1 and R^2 are different from each other and one of R^1 and R^2 is hydrogen and the other is selected from the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$, $-CH_2-O-PO(OH)_2$, and $-CH_2OH$, and whereby A stands for $-CH_2OH$ or -CHO.

34. The optionally isotope-labelled compound according to claim 33, wherein A stands for $-CH_2OH$.

35. The optionally isotope-labelled compound according to claim 33, wherein A stands for -CHO.

36. The optionally isotope-labelled compound according to one of claims 33 to 35, wherein R^1 is H and R^2 is selected from the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$ and $-CH_2-O-PO(OH)_2$.

37. The optionally isotope-labelled compound according to one of claims 33 to 35, wherein R^2 is H and R^1 is selected from the group consisting of $-CH_2-O-PO(OH)-O-PO(OH)_2$ and $-CH_2-O-PO(OH)_2$.

38. The optionally isotope-labelled compound according to one of claims 33 to 37, whereby said group consists of $-CH_2-O-PO(OH)-O-PO(OH)_2$.

39. Optionally isotope-labelled 1-hydroxy-2-methyl-2-butenyl 4-diphosphate salt or a protonated form thereof.

40. Optionally isotope-labelled (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate salt or a protonated form thereof.

41. Optionally isotope-labelled (*Z*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate salt or a protonated form thereof.

42. Use of a compound according to one of claims 33 to 41, notably according to claim 38, for screening for genes, enzymes or inhibitors of the biosynthesis of isoprenoids or terpenoids, either *in vitro* in the presence of an electron donor or *in vivo*.

43. Use of a compound according to one of claims 33 to 41, notably according to claim 40, as an immunomodulatory agent.

44. Use of a compound according to one of claims 33 to 41, notably according to claim 40, for activating γδ T cells.

45. Use of a compound according to one of claims 33 to 41, notably according to claim 40, for the preparation of a medicament.

46. Pharmaceutical composition containing a compound according to one of claims 33 to 41, notably according to claim 36 or 40, and a pharmaceutically acceptable carrier.

47. The pharmaceutical composition according to claim 46 further comprising an antibiotically active compound.

48. The pharmaceutical composition according to claim 47, wherein the antibiotically active compound is bacteriostatic.

49. The pharmaceutical composition according to claim 47, wherein the antibiotically active compound inhibits bacterial protein synthesis.

50. A method of treating a pathogen infection comprising administering a pharmaceutical composition according to one of claims 46 to 49.

51. Monoclonal or polyclonal antibody against a compound of one of claims 33 to 41.

52. A method of detecting a pathogen, notably in a body fluid, by using the antibody of claim 51.

53. Use of the cells, cell cultures or organisms or parts thereof in accordance with claims 15 to 24 for the production of a protein in an enzymatically competent form for the conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate.

54. Use of the cells, cell cultures or organisms or parts thereof in accordance with claims 15 to 24 for the production of a protein in an enzymatically competent form for the conversion of (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate into isopentenyl diphosphate and/or dimethylallyl diphosphate.

55. Use of the cells, cell cultures or organisms or parts thereof in accordance with claims 15 to 24 for the production of proteins in an enzymatically competent form for the conversion of 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate into isopentenyl diphosphate and/or dimethylallyl diphosphate.

56. A method of altering the expression level of the gene product(s) of *ispG* and/or *ispH* in cells comprising

(a) transforming host cells with the *ispG* and/or *ispH* gene,

(b) growing the transformed host cells of step (a) under conditions that are suitable for the efficient expression of *ispG* and/or *ispH*, resulting in production of altered levels of the ispG and/or ispH gene product(s) in the transformed cells relative to expression levels of untransformed cells.

57. Method of identifying an inhibitior of an enzyme functional for the conversion of 2C-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably its E-form, of the non-mevalonate isoprenoid pathway by the following steps:

(a) incubating a mixture containing said enzyme with its, optionally isotope-labeled, substrate 2C-methyl-D-erythritol-2,4-cyclodiphosphate under conditions suitable for said conversion in the presence and in the absence of a potential inhibitor,

(b) subsequently determining the concentration of 2C-methyl-D-erythritol 2,4-cyclodiphosphate and/or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, and

(c) comparing the concentration in the presence and in the absence of said potential inhibitor.

58. Method of identifying an inhibitior of an enzyme functional for the conversion of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably its E-form, to isopentenyl diphosphate or dimethylallyl diphosphate of the non-mevalonate isoprenoid pathway by the following steps:

(a) incubating a mixture containing said enzyme with its, optionally isotope-labeled, substrate 1-hydroxy-2-methyl-2-butenyl 4-diphosphate under conditions suitable for said conversion in the presence and in the absence of a potential inhibitor,

(b) determining the concentration of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate or dimethylallyl diphosphate, and

(c) comparing the concentration in the presence and in the absence of said potential inhibitor.

59. The method according to claim 58, wherein step (a) is carried out in the presence of FAD.

60. The method of one of claims 57 to 59, which further comprises preparing cells recombinantly endowed with a gene coding for said enzyme, culturing said cells, preparing a crude extract of said cells, and using said crude extract in step (a).

61. The method according to one of claims 57 to 60, wherein said enzyme is a plant enzyme.

62. The method according to one of claims 57 to 60, wherein said enzyme is an enzyme of *Plasmodium falciparum*.

63. The method according to one of claims 57 to 60, wherein said enzyme is a bacterial enzyme.

64. The method according to one of claims 57 to 60, wherein the incubation of step (a) is carried out in the presence of a sulfhydryl reductant e.g. dithiothreitol.

65. The method according to one of claims 57 to 64, wherein the incubation in step (a) is carried out in the presence of a phosphatase inhibitor.

66. The method according to claim 65, wherein the phosphatase inhibitor is an alkali fluoride.

67. The method according to one of claims 57 to 66, wherein the incubation of step (a) is carried out in the presence of NADH or NADPH.

68. The method according to one of claims 57 to 67, wherein the incubation in step (a) is carried out in the presence of an inhibitor of an enzyme acting downstream of isopentenyl diphosphate or dimethylallyl diphosphate.

69. The method according to one of claims 57 to 68, wherein the incubation of step (a) is carried out in the presence of a salt selected from the group of Co^{2+}, Mn^{2+}, Fe^{2+}, Ni^{2+} salts.

70. The method according to one of claims 57 to 69, wherein step (b) is carried out by reversed phase ion-pair HPLC chromatography.

71. The method according to one of claims 57 to 69, wherein step (b) is carried out by determining the consumption of NADH or NADPH.

72. The method according to one of claims 57 to 71, which is carried out on many potential inhibitors simultaneously or consecutively in a high-throughput screening.

73. A process for the efficient *in vivo* synthesis of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate; or isopentenyl

diphosphate or dimethylallyl diphosphate; optionally isotope labelled, in salt form or in protonated form, by the following steps:

(a) culturing cells, preferably bacterial cells, recombinantly endowed in accordance with one of claims 15 to 24 for said synthesis for a predetermined period of time at a predetermined temperature;

(b) optionally adding glucose to a predetermined final concentration and further culturing for a predetermined period of time;

(c) harvesting the cells;

(d) preparing a crude extract from the harvested cells;

(e) separating and purifying optionally isotope-labelled 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate; or isopentenyl diphosphate; or dimethylallyl diphosphate; in salt form or in protonated form, optionally by preparative chromatography.

74. A process for screening chemical libraries for the presence or absence of inhibition of the biosynthesis of isoprenoids, notably by blocking the biosynthesis of the intermediates 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate and/or dimethylallyl diphosphate, said screening comprising:

(a) culturing cells, preferably bacterial cells, recombinantly endowed in accordance with claim 15 for a predetermined period of time at a predetermined temperature;

(b) optionally adding glucose to a predetermined final concentration and further culturing for a predetermined period of time;

(c) harvesting the cells;

(d) preparing a crude extract from the harvested cells;

whereby steps (a) to (d) are carried out in the presence and in the absence of a prospective inhibitor;

(e) detecting difference(s) in the level(s) of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate and/or dimethylallyl diphosphate; between the presence and absence of the prospective inhibitor; and

(f) correlating said detected difference(s) with the presence or absence of an above-defined inhibition.

75. Cells, cell cultures or organisms or parts thereof for the efficient formation of a biosynthetic product or intermediate of the non-mevalonate pathway to isoprenoids or terpenoids, characterized by

(a) first recombinant endowment with a gene functional for the biosynthesis of 1-deoxy-D-xylulose 5-phosphate from 1-deoxy-D-xylulose;

(b) capability for the uptake of 1-deoxy-D-xylulose; and

(c) recombinant endowment(s) with gene(s) being functional for the conversion of 1-deoxy-D-xylulose 5-phosphate into desired downstream C5-intermediate(s) of said pathway.

76. Cells, cell cultures or organisms or parts thereof in accordance with claim 75, wherein said gene(s) of said second recombinant endowment(s) code(s) for enzyme(s) for the formation of at least one of the following C5-intermediates of the non-mevalonate isoprenoid pathway:

(a) 2*C*-methyl-D-erythritol 4-phosphate;

(b) 4-diphosphocytidyl-2*C*-methyl-D-erythritol;

(c) 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate;

(d) 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate;

(e) 1-hydroxy-2-methyl-2-butenyl 4-diphosphate;

(f) isopentenyl diphosphate;

(g) dimethylallyl diphosphate.

77. Cells, cell cultures or organisms or parts thereof in accordance with claim 75 or 76, characterized by the recombinant endowment with sets of genes as follows:

(a) *xylB, dxr*; or

(b) *xylB, dxr, ispD* (formerly *ygbP*); or

(c) *xylB, dxr, ispD, ispE* (formerly *ychB*), or

(d) *xylB, dxr, ispD, ispE, ispF* (formerly *ygbB*); or

(e) *xylB, dxr, ispD, ispE, ispF, ispG* (formerly *gcpE*); or

(f) *xylB, dxr, ispD, ispE, ispF, ispG, ispH* (formerly *lytB)*

of *E. coli* or a function-conservative homologue thereof and/or a function-conservative fusion, deletion or insertion variant of any of the above genes.

78. Cells, cell cultures or organisms or parts thereof in accordance with claim 75, characterized by the recombinant endowment with *xylB* and *ispG* (formerly *gcpE*) and optionally at least one gene selected from the following group: *dxr*; *ispD* (formerly *ygbP*); *ispE* (formerly *ychB*); *ispF* (formerly *ygbB*) of *E. coli* or a function-conservative homologue thereof, or a function-conservative fusion, deletion or insertion variant of any of the above genes.

79. Cells, cell cultures or organisms or parts thereof in accordance with claim 75, characterized by the recombinant endowment with *xylB* and *ispH* (formerly *lytB*) and optionally at least one gene selected from the following group: *dxr*; *ispD* (formerly *ygbP*); *ispE* (formerly *ychB*); *ispF* (formerly *ygbB*); *ispG* (formerly *gcpE*) of *E. coli* or a function-conservative homologue thereof, or a function-conservative fusion, deletion or insertion variant of any of the above genes.

80. Cells, cell cultures or organisms or parts thereof in accordance with claim 75, characterized by the recombinant endowment with *xyl, ispG,* (formerly *gcpE*) and *ispH* (formerly *lytB*) and optionally at least one gene selected from the following group: *dxr*; *ispD* (formerly *ygbP*); *ispE* (formerly *ychB*); *ispF* (formerly *ygbB*); of *E. coli* or a function-conservative homologue thereof, or a function-conservative fusion, deletion or insertion variant of any of the above genes.

81. A process for the efficient *in vivo* synthesis of 2*C*-methyl-D-erythritol 4-phosphate; or 4-diphosphocytidyl-2*C*-methyl-D-erythritol; or 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate; or 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate; or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate; or isopentenyl diphosphate or dimethylallyl diphosphate; optionally isotope labelled, in salt form or in protonated form, by the following steps:

(a) culturing cells, preferably bacterial cells, recombinantly endowed in accordance with one of claims 75 to 80 for said synthesis for a predetermined period of time at a predetermined temperature;

(b) adding 1-deoxy-D-xylulose to a predetermined final concentration and further culturing for a predetermined period of time;

(c) harvesting the cells;

(d) preparing a crude extract from the harvested cells;

(e) separating and purifying optionally isotope-labelled 2*C*-methyl-D-erythritol 4-phosphate; or 4-diphosphocytidyl-2*C*-methyl-D-erythritol; or 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate; or 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate; or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate; or isopentenyl diphosphate; or dimethylallyl diphosphate; in salt form or in protonated form, by preparative chromatography.

82. The process according to one of claims 73, 74 or 81, wherein step (a) is carried out in terrific broth medium.

83. The process according to one of claims 73, 74, 81 or 82, wherein a source for CTP, preferably cytidine or uridine, is added in step (a).

84. The process according to one of claims 73, 74 or 81 to 83, wherein a source of phosphorylation activity, preferably glycerol 3-phosphate and/or inorganic phosphate, is added in step (a).

85. The process according to one of claims 73, 74 or 81 to 84 wherein a source of reduction equivalents, preferably succinate and/or lipids and/or glucose and/or glycerol and/or lactate is added in step (a).

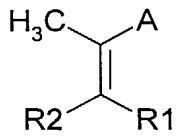
86. A process for screening chemical libraries for the presence or absence of inhibition of the biosynthesis of isoprenoids, notably by blocking the biosynthesis of the intermediates 2*C*-methyl-D-erythritol 4-phosphate and/or 4-diphosphocytidyl-2*C*-methyl-D-erythritol and/or 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate and/or 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and/or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate and/or dimethylallyl diphosphate, said screening comprising: (i) carrying out the steps (a) to (d) of claim 81, preferably in combination

with one of claims 82 to 85, in the presence and absence of a prospective inhibitor; (ii) detecting difference(s) in the level(s) of 1-deoxy-D-xylulose and/or 1-deoxy-D-xylulose 5-phosphate and/or 2C-methyl-D-erythritol 4-phosphate and/or 4-diphosphocytidyl-2*C*-methyl-D-erythritol and/or 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate and/or 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate and/or 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, notably (*E*)-1-hydroxy-2-methyl-2-butenyl 4-diphosphate and/or isopentenyl diphosphate and/or dimethylallyl diphosphate; between the presence and absence of a prospective inhibitor and (iii) correlating said detected difference(s) with the presence or absence of an above-defined inhibition.

87. The process according to claim 74 or 86, wherein said detecting of step (ii) is done by HPLC and/or NMR spectroscopy.

88. Vektor comprising a sequence coding for one of the recombinant endowments as defined in one of claims 75 to 80.

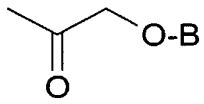
89. A process for the chemical preparation of a compound of formula I or a salt thereof:



(I)

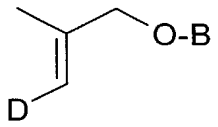
wherein A represents $-CH_2OH$ and R^1 and R^2 are different from each other and one of R^1 and R^2 is hydrogen and the other is $-CH_2-O-PO(OH)-O-PO(OH)_2$, $-CH_2-O-PO(OH)_2$ or $-CH_2-OH$ by the following steps:

(a) converting a compound of the following formula (II):

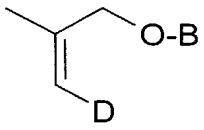


(II)

wherein B is a protective group into a compound of the following formula (III) or (IV):



(III)



(IV)

by a Wittig or Horner reagent, wherein the group D is a precursor group convertible reductively to a $-CH_2-OH$ group;

(b) reductively converting group D to a $-CH_2-OH$ group;

(c) optionally converting group $-CH_2-OH$ obtained in step (b) into $-CH_2-O-PO(OH)-O-PO(OH)_2$ or $-CH_2-O-PO(OH)_2$ or salts thereof in a manner knwon per se;

(d) optionally conversion to a desired salt;

(e) removing the protective group B.

90. The process according to claim 89, wherein said protective group B forms an acetal together with the remaining moiety of the compound of formula (II).

91. The process according to claim 89 or 90, wherein said protective group B is a 2-tetrahydropyranyl group.

92. The process according to one of claims 89 to 91, wherein group D is an alkoxycarbonyl group.

93. The process according to one of claims 89 to 92, wherein said reduction of step (b) is performed with a metal hydride, notably an aluminium hydride or a boron hydride.

94. The process according to one of claims 89 to 93, wherein step (c) comprises converting said $-CH_2-OH$ group to a $-CH_2-$halide group.

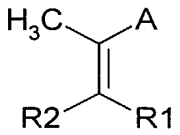
95. The process according to one of claims 89 to 94, wherein step (c) comprises reacting said $-CH_2-OH$ group with a sulfonic acid halogenide, notably tosyl chloride.

96. The process of one of claims 89 to 95, wherein step (c) comprises a reaction with phosphoric acid or diphosphoric acid or a salt thereof.

97. The process of one of claims 89 to 96, wherein steps (a) to (c) are carried out in aprotic solvents.

98. The process of one of claims 89 to 97, wherein step (e) is carried out by acid hydrolysis.

99. A process for the chemical preparation of a compound of formula I or a salt thereof:



(I)

wherein A represents $-CH_2OH$ or -CHO, R^1 is hydrogen, and R^2 is $-CH_2-O-PO(OH)-O-PO(OH)_2$, $-CH_2-O-PO(OH)_2$ or $-CH_2-OH$ by the following steps:

(a) converting 2-methyl-2-vinyl-oxiran into 4-chloro-2-methyl-2-buten-1-al;

(b) converting 4-chloro-2-methyl-2-buten-1-al to its acetal;

(c) substituting the chlorine atom in the product of step (b) by a hydroxyl group, a phosphate group or a pyrophosphate group;

(d) hydrolysing the acetal obtained in step (c) to produce an aldehyde group;

(e) optionally converting the aldehyde group of the product of step (d) to a $-CH_2OH$ group.

100. The process of claim 99, wherein step (a) is carried out in the presence of $CuCl_2$.

101. The process of claim 99 or 100, wherein step (b) is carried out in the presence of an ortho alkyl ester of formic acid.

102. The process of one of claims 99 to 101, wherein R^2 is $-CH_2-O-PO(OH)-O-PO(OH)_2$ or $-CH_2-O-PO(OH)_2$ and step (c) is carried out by reacting the product of step (b) with a tetra-alkylammonium pyrophosphate or a tetra-alkylammonium phosphate, respectively, in a polar aprotic solvent.

103. The process of one of claims 99 to 102, wherein step (e) is performed with an alkali metal borohydride in aqueous solution.

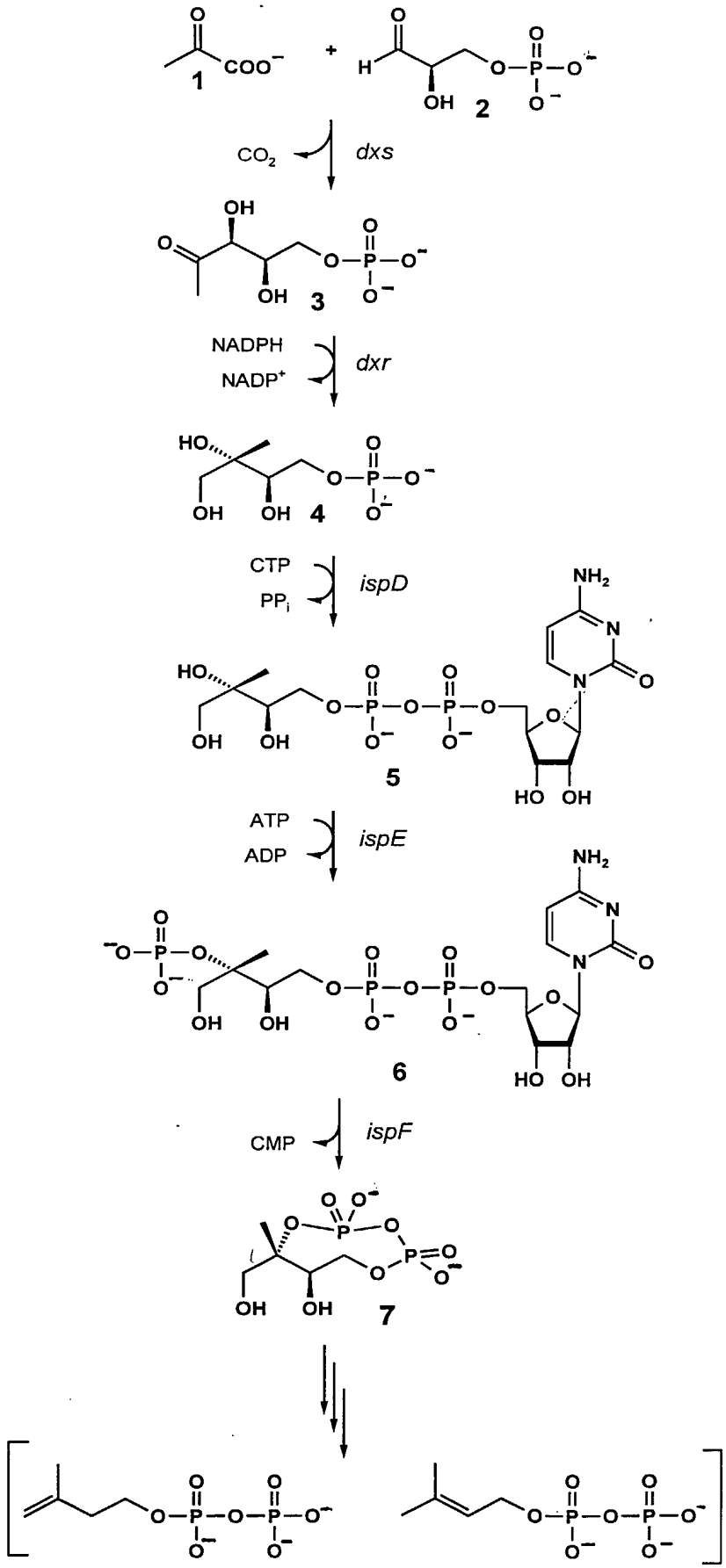

1
2
dxs
CO2
3
NADPH
NADP+
dxr
4
CTP
PPi
ispD
5
ATP
ADP
ispE
6
CMP
ispF
7

Fig. 1

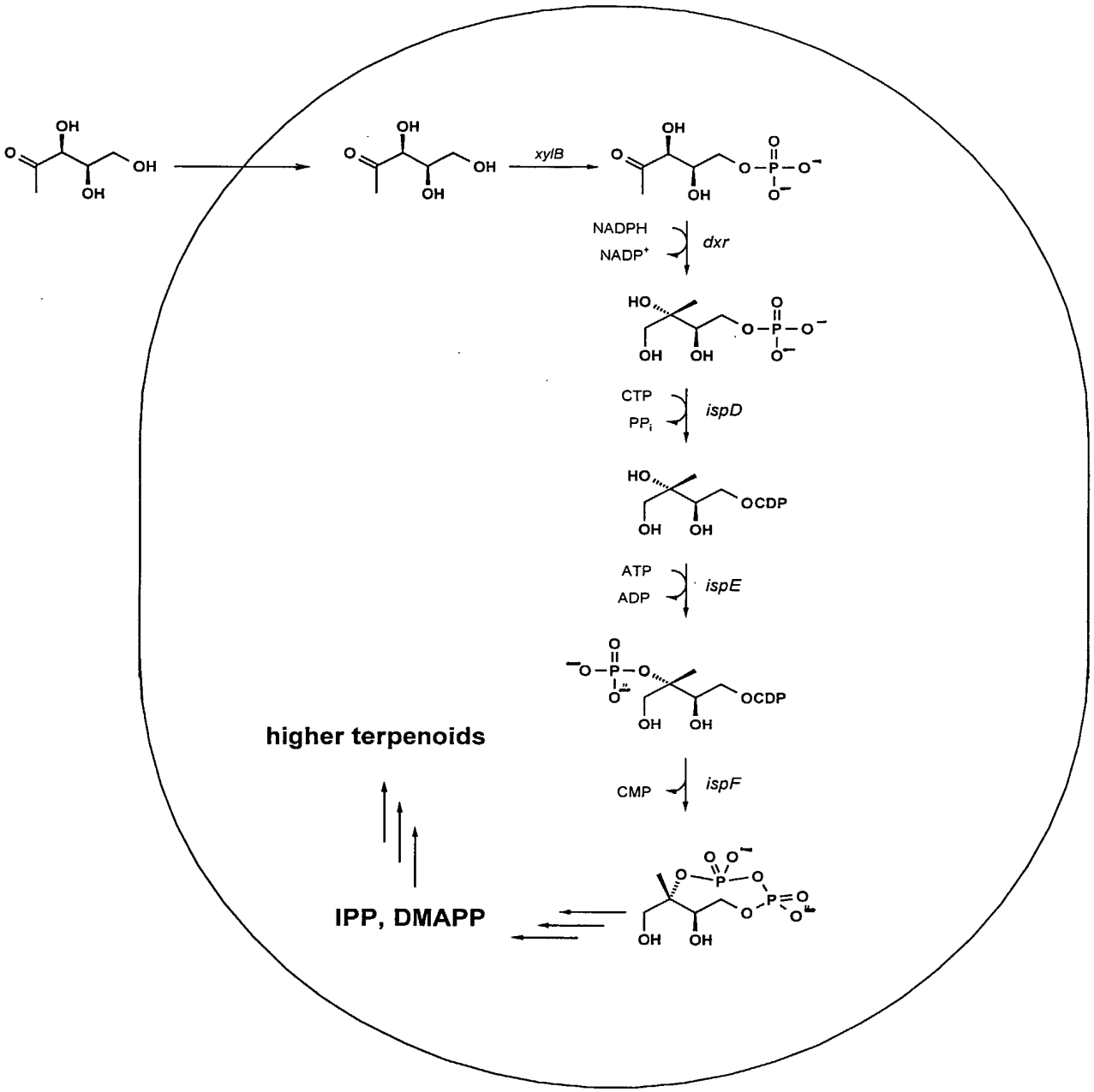

xylB
NADPH
NADP+
dxr
CTP
PPi
ispD
ATP
ADP
ispE
CMP
ispF
OCDP
higher terpenoids
IPP, DMAPP

Fig. 2

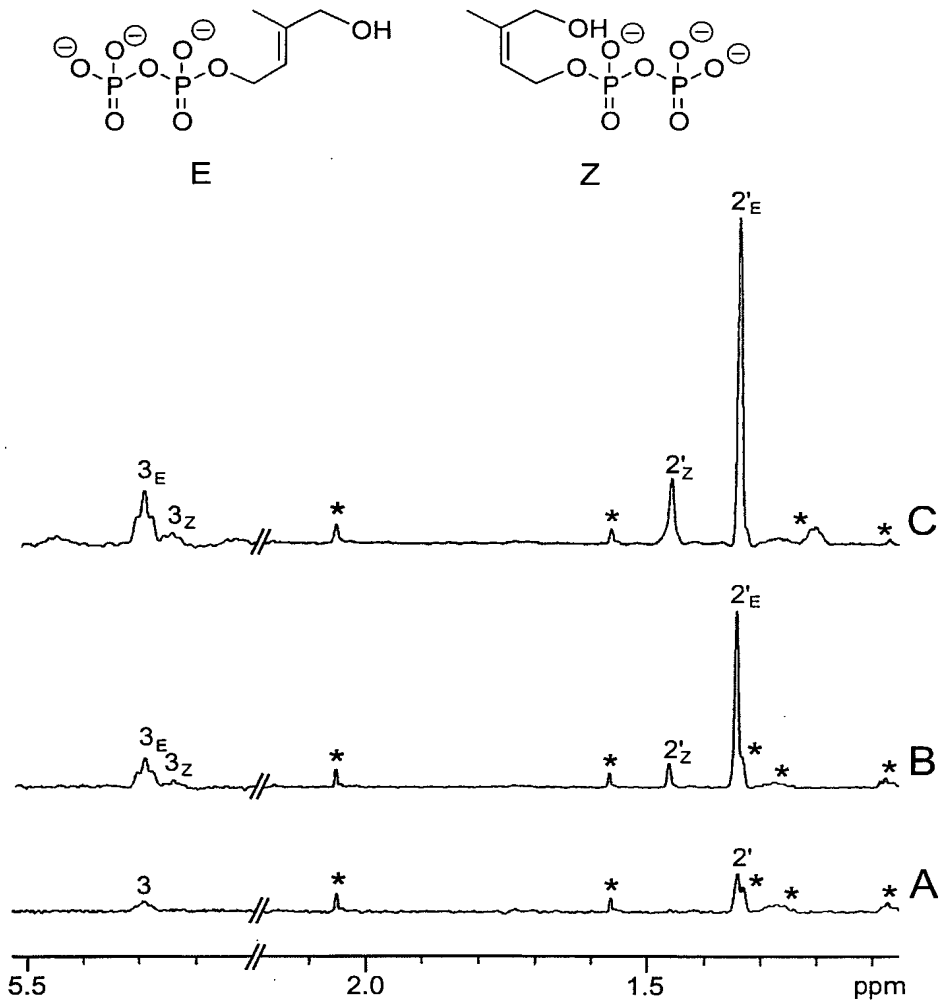


Fig. 3

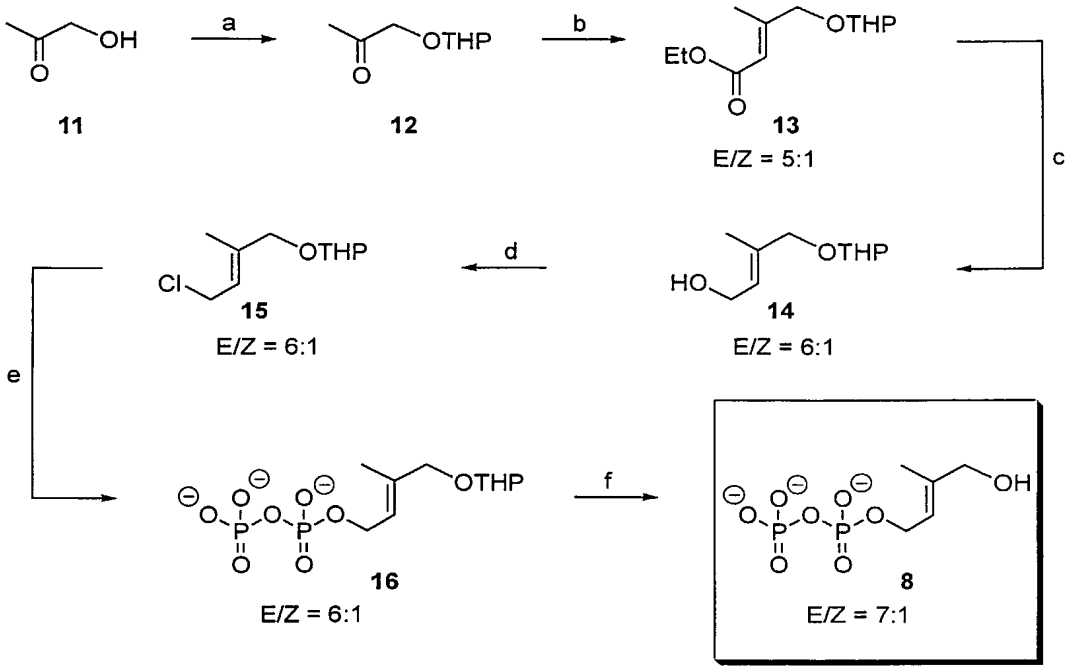

OH
a
OTHP
b
EtO
OTHP
11
12
13
E/Z = 5:1
c
OTHP
HO
14
E/Z = 6:1
d
OTHP
Cl
15
E/Z = 6:1
e
OTHP
f
16
E/Z = 6:1
OH
8
E/Z = 7:1

Fig. 4

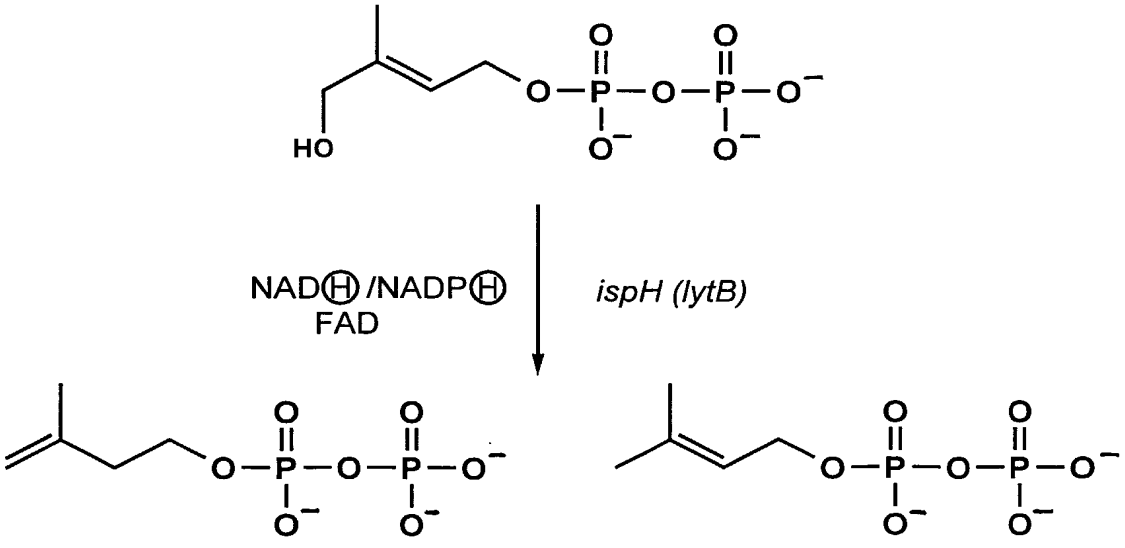

HO
NAD⊕H /NADP⊕H
FAD
ispH (lytB)

Fig. 5

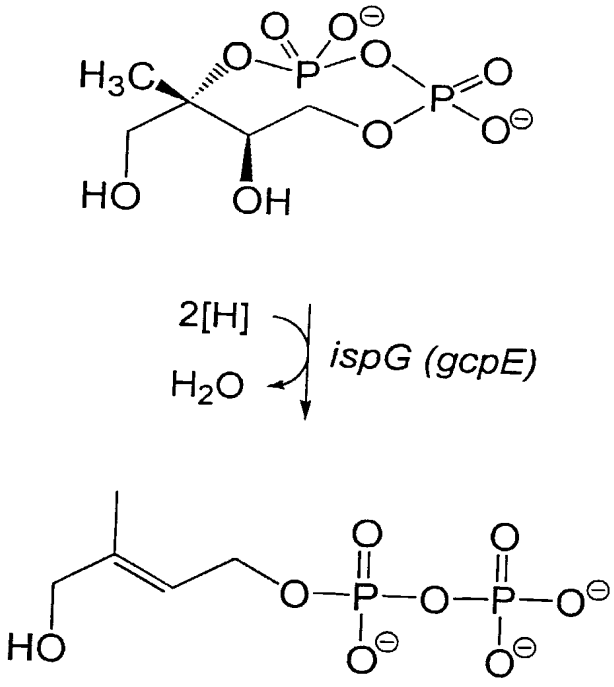

H3C
HO
OH
2[H]
H2O
ispG (gcpE)
HO

Fig. 6

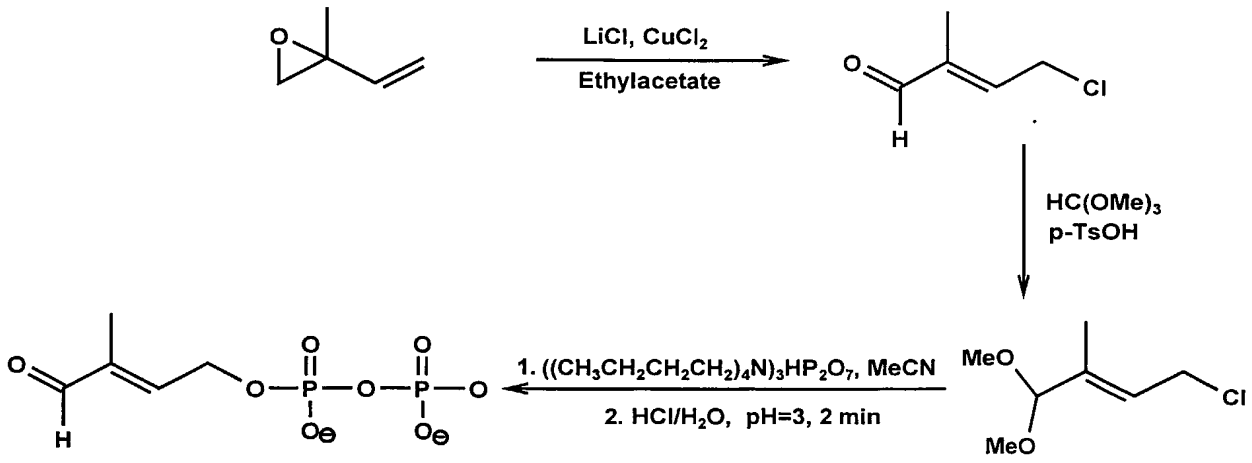

LiCl, CuCl2
Ethylacetate
Cl
HC(OMe)3
p-TsOH
MeO
MeO
Cl
1. ((CH3CH2CH2CH2)4N)3HP2O7, MeCN
2. HCl/H2O, pH=3, 2 min
H

Fig. 7

Annex A

DNA sequence of the vector construct pBSxylBdxr

ID PBSXYLBDXR PRELIMINARY; DNA; 5628 BP.

SQ SEQUENCE 5628 BP; 1378 A; 1374 C; 1552 G; 1324 T; 0 OTHER;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG

AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG
GCCGCTCTAG AACTAGTGGA TCCCCCGGGC TGCAGGAATT CGAGGAGAAA TTAACCATGT
ATATCGGGAT AGATCTTGGC ACCTCGGGCG TAAAAGTTAT TTTGCTCAAC GAGCAGGGTG
AGGTGGTTGC TGCGCAAACG GAAAAGCTGA CCGTTTCGCG CCCGCATCCA CTCTGGTCGG
AACAAGACCC GGAACAGTGG TGGCAGGCAA CTGATCGCGC AATGAAAGCT CTGGGCGATC
AGCATTCTCT GCAGGACGTT AAAGCATTGG GTATTGCCGG CCAGATGCAC GGAGCAACCT
TGCTGGATGC TCAGCAACGG GTGTTACGCC CTGCCATTTT GTGGAACGAC GGGCGCTGTG
CGCAAGAGTG CACTTTGCTG GAAGCGCGAG TTCCGCAATC GCGGGTGATT ACCGGCAACC
TGATGATGCC CGGATTTACT GCGCCTAAAT TGCTATGGGT TCAGCGGCAT GAGCCGGAGA
TATTCCGTCA AATCGACAAA GTATTATTAC CGAAAGATTA CTTGCGTCTG CGTATGACGG
GGGAGTTTGC CAGCGATATG TCTGACGCAG CTGGCACCAT GTGGCTGGAT GTCGCAAAGC
GTGACTGGAG TGACGTCATG CTGCAGGCTT GCGACTTATC TCGTGACCAG ATGCCCGCAT
TATACGAAGG CAGCGAAATT ACTGGTGCTT TGTTACCTGA AGTTGCGAAA GCGTGGGGTA
TGGCGACGGT GCCAGTTGTC GCAGGCGGTG GCGACAATGC AGCTGGTGCA GTTGGTGTGG
GAATGGTTGA TGCTAATCAG GCAATGTTAT CGCTGGGGAC GTCGGGGGTC TATTTTGCTG
TCAGCGAAGG GTTCTTAAGC AAGCCAGAAA GCGCCGTACA TAGCTTTTGC CATGCGCTAC
CGCAACGTTG GCATTTAATG TCTGTGATGC TGAGTGCAGC GTCGTGTCTG GATTGGGCCG
CGAAATTAAC CGGCCTGAGC AATGTCCCAG CTTTAATCGC TGCAGCTCAA CAGGCTGATG
AAAGTGCCGA GCCAGTTTGG TTTCTGCCTT ATCTTTCCGG CGAGCGTACG CCACACAATA
ATCCCCAGGC GAAGGGGGTT TTCTTTGGTT TGACTCATCA ACATGGCCCC AATGAACTGG
CGCGAGCAGT GCTGGAAGGC GTGGGTTATG CGCTGGCAGA TGGCATGGAT GTCGTGCATG
CCTGCGGTAT TAAACCGCAA AGTGTTACGT TGATTGGGGG CGGGGCGCGT AGTGAGTACT
GGCGTCAGAT GCTGGCGGAT ATCAGCGGTC AGCAGCTCGA TTACCGTACG GGGGGGGATG
TGGGGCCAGC ACTGGGCGCA GCAAGGCTGG CGCAGATCGC GGCGAATCCA GAGAAATCGC
TCATTGAATT GTTGCCGCAA CTACCGTTAG AACAGTCGCA TCTACCAGAT GCGCAGCGTT
ATGCCGCTTA TCAGCCACGA CGAGAAACGT TCCGTCGCCT CTATCAGCAA CTTCTGCCAT
TAATGGCGTA AAAGCTTGAG GAGAAATTAA CCATGAAGCA ACTCACCATT CTGGGCTCGA
CCGGCTCGAT TGGTTGCAGC ACGCTGGACG TGGTGCGCCA TAATCCCGAA CACTTCCGCG
TAGTTGCGCT GGTGGCAGGC AAAAATGTCA CTCGCATGGT AGAACAGTGC CTGGAATTCT
CTCCCCGCTA TGCCGTAATG GACGATGAAG CGAGTGCGAA ACTTCTTAAA ACGATGCTAC
AGCAACAGGG TAGCCGCACC GAAGTCTTAA GTGGGCAACA AGCCGCTTGC GATATGGCAG
CGCTTGAGGA TGTTGATCAG GTGATGGCAG CCATTGTTGG CGCTGCTGGG CTGTTACCTA

CGCTTGCTGC GATCCGCGCG GGTAAAACCA TTTTGCTGGC CAATAAAGAA TCACTGGTTA
CCTGCGGACG TCTGTTTATG GACGCCGTAA AGCAGAGCAA AGCGCAATTG TTACCGGTCG
ATAGCGAACA TAACGCCATT TTTCAGAGTT TACCGCAACC TATCCAGCAT AATCTGGGAT
ACGCTGACCT TGAGCAAAAT GGCGTGGTGT CCATTTTACT TACCGGGTCT GGTGGCCCTT
TCCGTGAGAC GCCATTGCGC GATTTGGCAA CAATGACGCC GGATCAAGCC TGCCGTCATC
CGAACTGGTC GATGGGGCGT AAAATTTCTG TCGATTCGGC TACCATGATG AACAAAGGTC
TGGAATACAT TGAAGCGCGT TGGCTGTTTA ACGCCAGCGC CAGCCAGATG GAAGTGCTGA
TTCACCCGCA GTCAGTGATT CACTCAATGG TGCGCTATCA GGACGGCAGT GTTCTGGCGC
AGCTGGGGGA ACCGGATATG CGTACGCCAA TTGCCCACAC CATGGCATGG CCGAATCGCG
TGAACTCTGG CGTGAAGCCG CTCGATTTTT GCAAACTAAG TGCGTTGACA TTTGCCGCAC
CGGATTATGA TCGTTATCCA TGCCTGAAAC TGGCGATGGA GGCGTTCGAA CAAGGCCAGG
CAGCGACGAC AGCATTGAAT GCCGCAAACG AAATCACCGT TGCTGCTTTT CTTGCGCAAC
AAATCCGCTT TACGGATATC GCTGCGTTGA ATTTATCCGT ACTGGAAAAA ATGGATATGC
GCGAACCACA ATGTGTGGAC GATGTGTTAT CTGTTGATGC GAACGCGCGT GAAGTCGCCA
GAAAAGAGGT GATGCGTCTC GCAAGCTGAG TCGACCTCGA GGGGGGGCCC GGTACCCAAT
TCGCCCTATA GTGAGTCGTA TTACGCGCGC TCACTGGCCG TCGTTTTACA ACGTCGTGAC
TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCCC TTTCGCCAGC
TGGCGTAATA GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT
GGCGAATGGA AATTGTAAGC GTTAATATTT TGTTAAAATT CGCGTTAAAT TTTTGTTAAA
TCAGCTCATT TTTTAACCAA TAGGCCGAAA TCGGCAAAAT CCCTTATAAA TCAAAAGAAT
AGACCGAGAT AGGGTTGAGT GTTGTTCCAG TTTGGAACAA GAGTCCACTA TTAAAGAACG
TGGACTCCAA CGTCAAAGGG CGAAAAACCG TCTATCAGGG CGATGGCCCA CTACGTGAAC
CATCACCCTA ATCAAGTTTT TTGGGGTCGA GGTGCCGTAA AGCACTAAAT CGGAACCCTA
AAGGGAGCCC CCGATTTAGA GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAAG
GGAAGAAAGC GAAAGGAGCG GGCGCTAGGG CGCTGGCAAG TGTAGCGGTC ACGCTGCGCG
TAACCACCAC ACCCGCCGCG CTTAATGCGC CGCTACAGGG CGCGTCAG

//

Annex B

DNA sequence of the vector construct pBSxylBdxrispD

ID PBSXYLBDXRISPD PRELIMINARY; DNA; 6354 BP.

SQ SEQUENCE 6354 BP; 1539 A; 1573 C; 1753 G; 1489 T;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG

AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG
GCCGCTCTAG AACTAGTGGA TCCCCCGGGC TGCAGGAATT CGAGGAGAAA TTAACCATGT
ATATCGGGAT AGATCTTGGC ACCTCGGGCG TAAAAGTTAT TTTGCTCAAC GAGCAGGGTG
AGGTGGTTGC TGCGCAAACG GAAAAGCTGA CCGTTTCGCG CCCGCATCCA CTCTGGTCGG
AACAAGACCC GGAACAGTGG TGGCAGGCAA CTGATCGCGC AATGAAAGCT CTGGGCGATC
AGCATTCTCT GCAGGACGTT AAAGCATTGG GTATTGCCGG CCAGATGCAC GGAGCAACCT
TGCTGGATGC TCAGCAACGG GTGTTACGCC CTGCCATTTT GTGGAACGAC GGGCGCTGTG
CGCAAGAGTG CACTTTGCTG GAAGCGCGAG TTCCGCAATC GCGGGTGATT ACCGGCAACC
TGATGATGCC CGGATTTACT GCGCCTAAAT TGCTATGGGT TCAGCGGCAT GAGCCGGAGA
TATTCCGTCA AATCGACAAA GTATTATTAC CGAAAGATTA CTTGCGTCTG CGTATGACGG
GGGAGTTTGC CAGCGATATG TCTGACGCAG CTGGCACCAT GTGGCTGGAT GTCGCAAAGC
GTGACTGGAG TGACGTCATG CTGCAGGCTT GCGACTTATC TCGTGACCAG ATGCCCGCAT
TATACGAAGG CAGCGAAATT ACTGGTGCTT TGTTACCTGA AGTTGCGAAA GCGTGGGGTA
TGGCGACGGT GCCAGTTGTC GCAGGCGGTG GCGACAATGC AGCTGGTGCA GTTGGTGTGG
GAATGGTTGA TGCTAATCAG GCAATGTTAT CGCTGGGGAC GTCGGGGGTC TATTTTGCTG
TCAGCGAAGG GTTCTTAAGC AAGCCAGAAA GCGCCGTACA TAGCTTTTGC CATGCGCTAC
CGCAACGTTG GCATTTAATG TCTGTGATGC TGAGTGCAGC GTCGTGTCTG GATTGGGCCG
CGAAATTAAC CGGCCTGAGC AATGTCCCAG CTTTAATCGC TGCAGCTCAA CAGGCTGATG
AAAGTGCCGA GCCAGTTTGG TTTCTGCCTT ATCTTTCCGG CGAGCGTACG CCACACAATA
ATCCCCAGGC GAAGGGGGTT TTCTTTGGTT TGACTCATCA ACATGGCCCC AATGAACTGG
CGCGAGCAGT GCTGGAAGGC GTGGGTTATG CGCTGGCAGA TGGCATGGAT GTCGTGCATG
CCTGCGGTAT TAAACCGCAA AGTGTTACGT TGATTGGGGG CGGGGCGCGT AGTGAGTACT
GGCGTCAGAT GCTGGCGGAT ATCAGCGGTC AGCAGCTCGA TTACCGTACG GGGGGGGATG
TGGGGCCAGC ACTGGGCGCA GCAAGGCTGG CGCAGATCGC GGCGAATCCA GAGAAATCGC
TCATTGAATT GTTGCCGCAA CTACCGTTAG AACAGTCGCA TCTACCAGAT GCGCAGCGTT
ATGCCGCTTA TCAGCCACGA CGAGAAACGT TCCGTCGCCT CTATCAGCAA CTTCTGCCAT
TAATGGCGTA AAAGCTTGAG GAGAAATTAA CCATGAAGCA ACTCACCATT CTGGGCTCGA
CCGGCTCGAT TGGTTGCAGC ACGCTGGACG TGGTGCGCCA TAATCCCGAA CACTTCCGCG
TAGTTGCGCT GGTGGCAGGC AAAAATGTCA CTCGCATGGT AGAACAGTGC CTGGAATTCT
CTCCCCGCTA TGCCGTAATG GACGATGAAG CGAGTGCGAA ACTTCTTAAA ACGATGCTAC
AGCAACAGGG TAGCCGCACC GAAGTCTTAA GTGGGCAACA AGCCGCTTGC GATATGGCAG
CGCTTGAGGA TGTTGATCAG GTGATGGCAG CCATTGTTGG CGCTGCTGGG CTGTTACCTA
CGCTTGCTGC GATCCGCGCG GGTAAAACCA TTTTGCTGGC CAATAAAGAA TCACTGGTTA

CCTGCGGACG TCTGTTTATG GACGCCGTAA AGCAGAGCAA AGCGCAATTG TTACCGGTCG
ATAGCGAACA TAACGCCATT TTTCAGAGTT TACCGCAACC TATCCAGCAT AATCTGGGAT
ACGCTGACCT TGAGCAAAAT GGCGTGGTGT CCATTTTACT TACCGGGTCT GGTGGCCCTT
TCCGTGAGAC GCCATTGCGC GATTTGGCAA CAATGACGCC GGATCAAGCC TGCCGTCATC
CGAACTGGTC GATGGGGCGT AAAATTTCTG TCGATTCGGC TACCATGATG AACAAAGGTC
TGGAATACAT TGAAGCGCGT TGGCTGTTTA ACGCCAGCGC CAGCCAGATG GAAGTGCTGA
TTCACCCGCA GTCAGTGATT CACTCAATGG TGCGCTATCA GGACGGCAGT GTTCTGGCGC
AGCTGGGGGA ACCGGATATG CGTACGCCAA TTGCCCACAC CATGGCATGG CCGAATCGCG
TGAACTCTGG CGTGAAGCCG CTCGATTTTT GCAAACTAAG TGCGTTGACA TTTGCCGCAC
CGGATTATGA TCGTTATCCA TGCCTGAAAC TGGCGATGGA GGCGTTCGAA CAAGGCCAGG
CAGCGACGAC AGCATTGAAT GCCGCAAACG AAATCACCGT TGCTGCTTTT CTTGCGCAAC
AAATCCGCTT TACGGATATC GCTGCGTTGA ATTTATCCGT ACTGGAAAAA ATGGATATGC
GCGAACCACA ATGTGTGGAC GATGTGTTAT CTGTTGATGC GAACGCGCGT GAAGTCGCCA
GAAAAGAGGT GATGCGTCTC GCAAGCTGAG TCGACGAGGA GAAATTAACC ATGGCAACCA
CTCATTTGGA TGTTTGCGCC GTGGTTCCGG CGGCCGGATT TGGCCGTCGA ATGCAAACGG
AATGTCCTAA GCAATATCTC TCAATCGGTA ATCAAACCAT TCTTGAACAC TCGGTGCATG
CGCTGCTGGC GCATCCCCGG GTGAAACGTG TCGTCATTGC CATAAGTCCT GGCGATAGCC
GTTTTGCACA ACTTCCTCTG GCGAATCATC CGCAAATCAC CGTTGTAGAT GGCGGTGATG
AGCGTGCCGA TTCCGTGCTG GCAGGTCTGA AAGCCGCTGG CGACGCGCAG TGGGTATTGG
TGCATGACGC CGCTCGTCCT TGTTTGCATC AGGATGACCT CGCGCGATTG TTGGCGTTGA
GCGAAACCAG CCGCACGGGG GGGATCCTCG CCGCACCAGT GCGCGATACT ATGAAACGTG
CCGAACCGGG CAAAAATGCC ATTGCTCATA CCGTTGATCG CAACGGCTTA TGGCACGCGC
TGACGCCGCA ATTTTTCCCT CGTGAGCTGT TACATGACTG TCTGACGCGC GCTCTAAATG
AAGGCGCGAC TATTACCGAC GAAGCCTCGG CGCTGGAATA TTGCGGATTC CATCCTCAGT
TGGTCGAAGG CCGTGCGGAT AACATTAAAG TCACGCGCCC GGAAGATTTG GCACTGGCCG
AGTTTTACCT CACCCGAACC ATCCATCAGG AGAATACATA ACTCGAGGGG GGGCCCGGTA
CCCAATTCGC CCTATAGTGA GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT
CGTGACTGGG AAAACCCTGG CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCTTTC
GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC CTTCCCAACA GTTGCGCAGC
CTGAATGGCG AATGGAAATT GTAAGCGTTA ATATTTTGTT AAAATTCGCG TTAAATTTTT
GTTAAATCAG CTCATTTTTT AACCAATAGG CCGAAATCGG CAAAATCCCT TATAAATCAA
AAGAATAGAC CGAGATAGGG TTGAGTGTTG TTCCAGTTTG GAACAAGAGT CCACTATTAA
AGAACGTGGA CTCCAACGTC AAAGGGCGAA AAACCGTCTA TCAGGGCGAT GGCCCACTAC
GTGAACCATC ACCCTAATCA AGTTTTTTGG GGTCGAGGTG CCGTAAAGCA CTAAATCGGA
ACCCTAAAGG GAGCCCCCGA TTTAGAGCTT GACGGGGAAA GCCGGCGAAC GTGGCGAGAA
AGGAAGGGAA GAAAGCGAAA GGAGCGGGCG CTAGGGCGCT GGCAAGTGTA GCGGTCACGC
TGCGCGTAAC CACCACACCC GCCGCGCTTA ATGCGCCGCT ACAGGGCGCG TCAG

Annex C

DNA sequence of the vector construct pBScyclo

ID PBSCYCLO PRELIMINARY; DNA; 7691 BP.

SQ SEQUENCE 7691 BP; 1844 A; 1888 C; 2148 G; 1811 T;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG

AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG
GCCGCTCTAG AACTAGTGGA TCCCCCGGGC TGCAGGAATT CGAGGAGAAA TTAACCATGT
ATATCGGGAT AGATCTTGGC ACCTCGGGCG TAAAAGTTAT TTTGCTCAAC GAGCAGGGTG
AGGTGGTTGC TGCGCAAACG GAAAAGCTGA CCGTTTCGCG CCCGCATCCA CTCTGGTCGG
AACAAGACCC GGAACAGTGG TGGCAGGCAA CTGATCGCGC AATGAAAGCT CTGGGCGATC
AGCATTCTCT GCAGGACGTT AAAGCATTGG GTATTGCCGG CCAGATGCAC GGAGCAACCT
TGCTGGATGC TCAGCAACGG GTGTTACGCC CTGCCATTTT GTGGAACGAC GGGCGCTGTG
CGCAAGAGTG CACTTTGCTG GAAGCGCGAG TTCCGCAATC GCGGGTGATT ACCGGCAACC
TGATGATGCC CGGATTTACT GCGCCTAAAT TGCTATGGGT TCAGCGGCAT GAGCCGGAGA
TATTCCGTCA AATCGACAAA GTATTATTAC CGAAAGATTA CTTGCGTCTG CGTATGACGG
GGGAGTTTGC CAGCGATATG TCTGACGCAG CTGGCACCAT GTGGCTGGAT GTCGCAAAGC
GTGACTGGAG TGACGTCATG CTGCAGGCTT GCGACTTATC TCGTGACCAG ATGCCCGCAT
TATACGAAGG CAGCGAAATT ACTGGTGCTT TGTTACCTGA AGTTGCGAAA GCGTGGGGTA
TGGCGACGGT GCCAGTTGTC GCAGGCGGTG GCGACAATGC AGCTGGTGCA GTTGGTGTGG
GAATGGTTGA TGCTAATCAG GCAATGTTAT CGCTGGGGAC GTCGGGGGTC TATTTTGCTG
TCAGCGAAGG GTTCTTAAGC AAGCCAGAAA GCGCCGTACA TAGCTTTTGC CATGCGCTAC
CGCAACGTTG GCATTTAATG TCTGTGATGC TGAGTGCAGC GTCGTGTCTG GATTGGGCCG
CGAAATTAAC CGGCCTGAGC AATGTCCCAG CTTTAATCGC TGCAGCTCAA CAGGCTGATG
AAAGTGCCGA GCCAGTTTGG TTTCTGCCTT ATCTTTCCGG CGAGCGTACG CCACACAATA
ATCCCCAGGC GAAGGGGGTT TTCTTTGGTT TGACTCATCA ACATGGCCCC AATGAACTGG
CGCGAGCAGT GCTGGAAGGC GTGGGTTATG CGCTGGCAGA TGGCATGGAT GTCGTGCATG
CCTGCGGTAT TAAACCGCAA AGTGTTACGT TGATTGGGGG CGGGGCGCGT AGTGAGTACT
GGCGTCAGAT GCTGGCGGAT ATCAGCGGTC AGCAGCTCGA TTACCGTACG GGGGGGGATG
TGGGGCCAGC ACTGGGCGCA GCAAGGCTGG CGCAGATCGC GGCGAATCCA GAGAAATCGC
TCATTGAATT GTTGCCGCAA CTACCGTTAG AACAGTCGCA TCTACCAGAT GCGCAGCGTT
ATGCCGCTTA TCAGCCACGA CGAGAAACGT TCCGTCGCCT CTATCAGCAA CTTCTGCCAT
TAATGGCGTA AAAGCTTGAG GAGAAATTAA CCATGAAGCA ACTCACCATT CTGGGCTCGA
CCGGCTCGAT TGGTTGCAGC ACGCTGGACG TGGTGCGCCA TAATCCCGAA CACTTCCGCG
TAGTTGCGCT GGTGGCAGGC AAAAATGTCA CTCGCATGGT AGAACAGTGC CTGGAATTCT
CTCCCCGCTA TGCCGTAATG GACGATGAAG CGAGTGCGAA ACTTCTTAAA ACGATGCTAC
AGCAACAGGG TAGCCGCACC GAAGTCTTAA GTGGGCAACA AGCCGCTTGC GATATGGCAG
CGCTTGAGGA TGTTGATCAG GTGATGGCAG CCATTGTTGG CGCTGCTGGG CTGTTACCTA
CGCTTGCTGC GATCCGCGCG GGTAAAACCA TTTTGCTGGC CAATAAAGAA TCACTGGTTA

CCTGCGGACG TCTGTTTATG GACGCCGTAA AGCAGAGCAA AGCGCAATTG TTACCGGTCG
ATAGCGAACA TAACGCCATT TTTCAGAGTT TACCGCAACC TATCCAGCAT AATCTGGGAT
ACGCTGACCT TGAGCAAAAT GGCGTGGTGT CCATTTTACT TACCGGGTCT GGTGGCCCTT
TCCGTGAGAC GCCATTGCGC GATTTGGCAA CAATGACGCC GGATCAAGCC TGCCGTCATC
CGAACTGGTC GATGGGGCGT AAAATTTCTG TCGATTCGGC TACCATGATG AACAAAGGTC
TGGAATACAT TGAAGCGCGT TGGCTGTTTA ACGCCAGCGC CAGCCAGATG GAAGTGCTGA
TTCACCCGCA GTCAGTGATT CACTCAATGG TGCGCTATCA GGACGGCAGT GTTCTGGCGC
AGCTGGGGGA ACCGGATATG CGTACGCCAA TTGCCCACAC CATGGCATGG CCGAATCGCG
TGAACTCTGG CGTGAAGCCG CTCGATTTTT GCAAACTAAG TGCGTTGACA TTTGCCGCAC
CGGATTATGA TCGTTATCCA TGCCTGAAAC TGGCGATGGA GGCGTTCGAA CAAGGCCAGG
CAGCGACGAC AGCATTGAAT GCCGCAAACG AAATCACCGT TGCTGCTTTT CTTGCGCAAC
AAATCCGCTT TACGGATATC GCTGCGTTGA ATTTATCCGT ACTGGAAAAA ATGGATATGC
GCGAACCACA ATGTGTGGAC GATGTGTTAT CTGTTGATGC GAACGCGCGT GAAGTCGCCA
GAAAAGAGGT GATGCGTCTC GCAAGCTGAG TCGACGAGGA GAAATTAACC ATGGCAACCA
CTCATTTGGA TGTTTGCGCC GTGGTTCCGG CGGCCGGATT TGGCCGTCGA ATGCAAACGG
AATGTCCTAA GCAATATCTC TCAATCGGTA ATCAAACCAT TCTTGAACAC TCGGTGCATG
CGCTGCTGGC GCATCCCCGG GTGAAACGTG TCGTCATTGC CATAAGTCCT GGCGATAGCC
GTTTTGCACA ACTTCCTCTG GCGAATCATC CGCAAATCAC CGTTGTAGAT GGCGGTGATG
AGCGTGCCGA TTCCGTGCTG GCAGGTCTGA AAGCCGCTGG CGACGCGCAG TGGGTATTGG
TGCATGACGC CGCTCGTCCT TGTTTGCATC AGGATGACCT CGCGCGATTG TTGGCGTTGA
GCGAAACCAG CCGCACGGGG GGGATCCTCG CCGCACCAGT GCGCGATACT ATGAAACGTG
CCGAACCGGG CAAAAATGCC ATTGCTCATA CCGTTGATCG CAACGGCTTA TGGCACGCGC
TGACGCCGCA ATTTTTCCCT CGTGAGCTGT TACATGACTG TCTGACGCGC GCTCTAAATG
AAGGCGCGAC TATTACCGAC GAAGCCTCGG CGCTGGAATA TTGCGGATTC CATCCTCAGT
TGGTCGAAGG CCGTGCGGAT AACATTAAAG TCACGCGCCC GGAAGATTTG GCACTGGCCG
AGTTTTACCT CACCCGAACC ATCCATCAGG AGAATACATA ATGCGAATTG GACACGGTTT
TGACGTACAT GCCTTTGGCG GTGAAGGCCC AATTATCATT GGTGGCGTAC GCATTCCTTA
CGAAAAAGGA TTGCTGGCGC ATTCTGATGG CGACGTGGCG CTCCATGCGT TGACCGATGC
ATTGCTTGGC GCGGCGGCGC TGGGGGATAT CGGCAAGCTG TTCCCGGATA CCGATCCGGC
ATTTAAAGGT GCCGATAGCC GCGAGCTGCT ACGCGAAGCC TGGCGTCGTA TTCAGGCGAA
GGGTTATACC CTTGGCAACG TCGATGTCAC TATCATCGCT CAGGCACCGA AGATGTTGCC
GCACATTCCA CAAATGCGCG TGTTTATTGC CGAAGATCTC GGCTGCCATA TGGATGATGT
TAACGTGAAA GCCACTACTA CGGAAAAACT GGGATTTACC GGACGTGGGG AAGGGATTGC
CTGTGAAGCG GTGGCGCTAC TCATTAAGGC AACAAAATGA CTCGAGGAGG AGAAATTAAC
CATGCGGACA CAGTGGCCCT CTCCGGCAAA ACTTAATCTG TTTTTATACA TTACCGGTCA
GCGTGCGGAT GGTTACCACA CGCTGCAAAC GCTGTTTCAG TTTCTTGATT ACGGCGACAC
CATCAGCATT GAGCTTCGTG ACGATGGGGA TATTCGTCTG TTAACGCCCG TTGAAGGCGT

GGAACATGAA GATAACCTGA TCGTTCGCGC AGCGCGATTG TTGATGAAAA CTGCGGCAGA
CAGCGGGCGT CTTCCGACGG GAAGCGGTGC GAATATCAGC ATTGACAAGC GTTTGCCGAT
GGGCGGCGGT CTCGGCGGTG GTTCATCCAA TGCCGCGACG GTCCTGGTGG CATTAAATCA
TCTCTGGCAA TGCGGGCTAA GCATGGATGA GCTGGCGGAA ATGGGGCTGA CGCTGGGCGC
AGATGTTCCT GTCTTTGTTC GGGGGCATGC CGCGTTTGCC GAAGGCGTTG GTGAAATACT
AACGCCGGTG GATCCGCCAG AGAAGTGGTA TCTGGTGGCG CACCCTGGTG TAAGTATTCC
GACTCCGGTG ATTTTTAAAG ATCCTGAACT CCCGCGCAAT ACGCCAAAAA GGTCAATAGA
AACGTTGCTA AAATGTGAAT TCAGCAATGA TTGCGAGGTT ATCGCAAGAA AACGTTTTCG
CGAGGTTGAT GCGGTGCTTT CCTGGCTGTT AGAATACGCC CCGTCGCGCC TGACTGGGAC
AGGGGCCTGT GTCTTTGCTG AATTTGATAC AGAGTCTGAA GCCCGCCAGG TGCTAGAGCA
AGCCCCGGAA TGGCTCAATG GCTTTGTGGC GAAAGGCGCT AATCTTTCCC CATTGCACAG
AGCCATGCTT TAAGGTACCC AATTCGCCCT ATAGTGAGTC GTATTACGCG CGCTCACTGG
CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG
CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC GATCGCCCTT
CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGAAATTGTA AGCGTTAATA TTTTGTTAAA
ATTCGCGTTA AATTTTTGTT AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA
AATCCCTTAT AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTTGTTC CAGTTTGGAA
CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA CCGTCTATCA
GGGCGATGGC CCACTACGTG AACCATCACC CTAATCAAGT TTTTTGGGGT CGAGGTGCCG
TAAAGCACTA AATCGGAACC CTAAAGGGAG CCCCCGATTT AGAGCTTGAC GGGGAAAGCC
GGCGAACGTG GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA GGGCGCTGGC
AAGTGTAGCG GTCACGCTGC GCGTAACCAC CACACCCGCC GCGCTTAATG CGCCGCTACA
GGGCGCGTCA G

//

Annex D

DNA sequence of the vector construct pACYCgcpE

ID PACYCGCPE PRELIMINARY; DNA; 5109 BP.

SQ SEQUENCE 5109 BP; 1194 A; 1365 C; 1324 G; 1226 T; 0 OTHER;

GAATTCCGGA TGAGCATTCA TCAGGCGGGC AAGAATGTGA ATAAAGGCCG GATAAAACTT
GTGCTTATTT TTCTTTACGG TCTTTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT
ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA
TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA
AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT GGTGAAAGTT
GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC
GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG TCACAGGTAT
TTATTCGGCG CAAAGTGCGT CGGGTGATGC TGCCAACTTA CTGATTTAGT GTATGATGGT
GTTTTTGAGG TGCTCCAGTG GCTTCTGTTT CTATCAGCTG TCCCTCCTGT TCAGCTACTG
ACGGGGTGGT GCGTAACGGC AAAAGCACCG CCGGACATCA GCGCTAGCGG AGTGTATACT
GGCTTACTAT GTTGGCACTG ATGAGGGTGT CAGTGAAGTG CTTCATGTGG CAGGAGAAAA
AAGGCTGCAC CGGTGCGTCA GCAGAATATG TGATACAGGA TATATTCCGC TTCCTCGCTC
ACTGACTCGC TACGCTCGGT CGTTCGACTG CGGCGAGCGG AAATGGCTTA CGAACGGGGC
GGAGATTTCC TGGAAGATGC CAGGAAGATA CTTAACAGGG AAGTGAGAGG GCCGCGGCAA
AGCCGTTTTT CCATAGGCTC CGCCCCCCTG ACAAGCATCA CGAAATCTGA CGCTCAAATC
AGTGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC GTTTCCCCCT GGCGGCTCCC
TCGTGCGCTC TCCTGTTCCT GCCTTTCGGT TTACCGGTGT CATTCCGCTG TTATGGCCGC
GTTTGTCTCA TTCCACGCCT GACACTCAGT TCCGGGTAGG CAGTTCGCTC CAAGCTGGAC
TGTATGCACG AACCCCCCGT TCAGTCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
GAGTCCAACC CGGAAAGACA TGCAAAAGCA CCACTGGCAG CAGCCACTGG TAATTGATTT
AGAGGAGTTA GTCTTGAAGT CATGCGCCGG TTAAGGCTAA ACTGAAAGGA CAAGTTTTGG
TGACTGCGCT CCTCCAAGCC AGTTACCTCG GTTCAAAGAG TTGGTAGCTC AGAGAACCTT
CGAAAAACCG CCCTGCAAGG CGGTTTTTTC GTTTTCAGAG CAAGAGATTA CGCGCAGACC
AAAACGATCT CAAGAAGATC ATCTTATTAA TCAGATAAAA TATTTCTAGA TTTCAGTGCA
ATTTATCTCT TCAAATGTAG CACCTGAAGT CAGCCCCATA CGATATAAGT TGTAATTCTC
ATGTTTGACA GCTTATCATC GATAAGCTTT AATGCGGTAG TTTATCACAG TTAAATTGCT
AACGCAGTCA GGCACCGTGT ATGAAATCTA ACAATGCGCT CATCGTCATC CTCGGCACCG
TCACCCTGGA TGCTGTAGGC ATAGGCTTGG TTATGCCGGT ACTGCCGGGC CTCTTGCGGG
ATATCGTCCA TTCCGACAGC ATCGCCAGTC ACTATGGCGT GCTGCTAGCG CTATATGCGT
TGATGCAATT TCTATGCGCA CCCGTTCTCG GAGCACTGTC CGACCGCTTT GGCCGCCGCC
CAGTCCTGCT CGCTTCGCTA CTTGGAGCCA CTATCGACTA CGCGATCATG GCGACCACAC
CCGTCCTGTG GATCCGAGGA GAAATTAACC ATGCATAACC AGGCTCCAAT TCAACGTAGA

AAATCAACAC GTATTTACGT TGGGAATGTG CCGATTGGCG ATGGTGCTCC CATCGCCGTA
CAGTCCATGA CCAATACGCG TACGACAGAC GTCGAAGCAA CGGTCAATCA AATCAAGGCG
CTGGAACGCG TTGGCGCTGA TATCGTCCGT GTATCCGTAC CGACGATGGA CGCGGCAGAA
GCGTTCAAAC TCATCAAACA GCAGGTTAAC GTGCCGCTGG TGGCTGACAT CCACTTCGAC
TATCGCATTG CGCTGAAAGT AGCGGAATAC GGCGTCGATT GTCTGCGTAT TAACCCTGGC
AATATCGGTA ATGAAGAGCG TATTCGCATG GTGGTTGACT GTGCGCGCGA TAAAAACATT
CCGATCCGTA TTGGCGTTAA CGCCGGATCG CTGGAAAAAG ATCTGCAAGA AAAGTATGGC
GAACCGACGC CGCAGGCGTT GCTGGAATCT GCCATGCGTC ATGTTGATCA TCTCGATCGC
CTGAACTTCG ATCAGTTCAA AGTCAGCGTG AAAGCGTCTG ACGTCTTCCT CGCTGTTGAG
TCTTATCGTT TGCTGGCAAA ACAGATCGAT CAGCCGTTGC ATCTGGGGAT CACCGAAGCC
GGTGGTGCGC GCAGCGGGGC AGTAAAATCC GCCATTGGTT TAGGTCTGCT GCTGTCTGAA
GGCATCGGCG ACACGCTGCG CGTATCGCTG GCGGCCGATC CGGTCGAAGA GATCAAAGTC
GGTTTCGATA TTTTGAAATC GCTGCGTATC CGTTCGCGAG GGATCAACTT CATCGCCTGC
CCGACCTGTT CGCGTCAGGA ATTTGATGTT ATCGGTACGG TTAACGCGCT GGAGCAACGC
CTGGAAGATA TCATCACTCC GATGGACGTT TCGATTATCG GCTGCGTGGT GAATGGCCCA
GGTGAGGCGC TGGTTTCTAC ACTCGGCGTC ACCGGCGGCA ACAAGAAAAG CGGCCTCTAT
GAAGATGGCG TGCGCAAAGA CCGTCTGGAC AACAACGATA TGATCGACCA GCTGGAAGCA
CGCATTCGTG CGAAAGCCAG TCAGCTGGAC GAAGCGCGTC GAATTGACGT TCAGCAGGTT
GAAAAATAAG TCGACCGATG CCCTTGAGAG CCTTCAACCC AGTCAGCTCC TTCCGGTGGG
CGCGGGGCAT GACTATCGTC GCCGCACTTA TGACTGTCTT CTTTATCATG CAACTCGTAG
GACAGGTGCC GGCAGCGCTC TGGGTCATTT TCGGCGAGGA CCGCTTTCGC TGGAGCGCGA
CGATGATCGG CCTGTCGCTT GCGGTATTCG GAATCTTGCA CGCCCTCGCT CAAGCCTTCG
TCACTGGTCC CGCCACCAAA CGTTTCGGCG AGAAGCAGGC CATTATCGCC GGCATGGCGG
CCGACGCGCT GGGCTACGTC TTGCTGGCGT TCGCGACGCG AGGCTGGATG GCCTTCCCCA
TTATGATTCT TCTCGCTTCC GGCGGCATCG GGATGCCCGC GTTGCAGGCC ATGCTGTCCA
GGCAGGTAGA TGACGACCAT CAGGGACAGC TTCAAGGATC GCTCGCGGCT CTTACCAGCC
TAACTTCGAT CACTGGACCG CTGATCGTCA CGGCGATTTA TGCCGCCTCG GCGAGCACAT
GGAACGGGTT GGCATGGATT GTAGGCGCCG CCCTATACCT TGTCTGCCTC CCCGCGTTGC
GTCGCGGTGC ATGGAGCCGG GCCACCTCGA CCTGAATGGA AGCCGGCGGC ACCTCGCTAA
CGGATTCACC ACTCCAAGAA TTGGAGCCAA TCAATTCTTG CGGAGAACTG TGAATGCGCA
AACCAACCCT TGGCAGAACA TATCCATCGC GTCCGCCATC TCCAGCAGCC GCACGCGGCG
CATCTCGGGC AGCGTTGGGT CCTGGCCACG GGTGCGCATG ATCGTGCTCC TGTCGTTGAG
GACCCGGCTA GGCTGGCGGG GTTGCCTTAC TGGTTAGCAG AATGAATCAC CGATACGCGA
GCGAACGTGA AGCGACTGCT GCTGCAAAAC GTCTGCGACC TGAGCAACAA CATGAATGGT
CTTCGGTTTC CGTGTTTCGT AAAGTCTGGA AACGCGGAAG TCCCCTACGT GCTGCTGAAG
TTGCCCGCAA CAGAGAGTGG AACCAACCGG TGATACCACG ATACTATGAC TGAGAGTCAA
CGCCATGAGC GGCCTCATTT CTTATTCTGA GTTACAACAG TCCGCACCGC TGTCCGGTAG

CTCCTTCCGG TGGGCGCGGG GCATGACTAT CGTCGCCGCA CTTATGACTG TCTTCTTTAT
CATGCAACTC GTAGGACAGG TGCCGGCAGC GCCCAACAGT CCCCCGGCCA CGGGGCCTGC
CACCATACCC ACGCCGAAAC AAGCGCCCTG CACCATTATG TTCCGGATCT GCATCGCAGG
ATGCTGCTGG CTACCCTGTG GAACACCTAC ATCTGTATTA ACGAAGCGCT AACCGTTTTT
ATCAGGCTCT GGGAGGCAGA ATAAATGATC ATATCGTCAA TTATTACCTC CACGGGGAGA
GCCTGAGCAA ACTGGCCTCA GGCATTTGAG AAGCACACGG TCACACTGCT TCCGGTAGTC
AATAAACCGG TAAACCAGCA ATAGACATAA GCGGCTATTT AACGACCCTG CCCTGAACCG
ACGACCGGGT CGAATTTGCT TTCGAATTTC TGCCATTCAT CCGCTTATTA TCACTTATTC
AGGCGTAGCA CCAGGCGTTT AAGGGCACCA ATAACTGCCT TAAAAAAATT ACGCCCCGCC
CTGCCACTCA TCGCAGTACT GTTGTAATTC ATTAAGCATT CTGCCGACAT GGAAGCCATC
ACAGACGGCA TGATGAACCT GAATCGCCAG CGGCATCAGC ACCTTGTCGC CTTGCGTATA
ATATTTGCCC ATGGTGAAAA CGGGGGCGAA GAAGTTGTCC ATATTGGCCA CGTTTAAATC
AAAACTGGTG AAACTCACCC AGGGATTGGC TGAGACGAAA AACATATTCT CAATAAACCC
TTTAGGGAAA TAGGCCAGGT TTTCACCGTA ACACGCCACA TCTTGCGAAT ATATGTGTAG
AAACTGCCGG AAATCGTCGT GGTATTCACT CCAGAGCGAT GAAAACGTTT CAGTTTGCTC
ATGGAAAACG GTGTAACAAG GGTGAACACT ATCCCATATC ACCAGCTCAC CGTCTTTCAT
TGCCATACG

//

Annex E

DNA sequence of the plasmid pBScaro14

ID PBSCARO14 PRELIMINARY; DNA; 7494 BP.SQ SEQUENCE 7494 BP;
1722 A; 1935 C; 2026 G; 1811 T;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG
GCCGCTCTAG AACTAGTGGA TCCCCCGGGC TGCAGGAATT GCCGTAAATG TATCCGTTTA
TAAGGACAGC CCGAATGACG GTCTGCGCAA AAAAACACGT TCATCTCACT CGCGATGCTG
CGGAGCAGTT ACTGGCTGAT ATTGATCGAC GCCTTGATCA GTTATTGCCC GTGGAGGGAG
AACGGGATGT TGTGGGTGCC GCGATGCGTG AAGGTGCGCT GGCACCGGGA AAACGTATTC
GCCCCATGTT GCTGTTGCTG ACCGCCCGCG ATCTGGGTTG CGCTGTCAGC CATGACGGAT
TACTGGATTT GGCCTGTGCG GTGGAAATGG TCCACGCGGC TTCGCTGATC CTTGACGATA
TGCCCTGCAT GGACGATGCG AAGCTGCGGC GCGGACGCCC TACCATTCAT TCTCATTACG
GAGAGCATGT GGCAATACTG GCGGCGGTTG CCTTGCTGAG TAAAGCCTTT GGCGTAATTG
CCGATGCAGA TGGCCTCACG CCGCTGGCAA AAAATCGGGC GGTTTCTGAA CTGTCAAACG
CCATCGGCAT GCAAGGATTG GTTCAGGGTC AGTTCAAGGA TCTGTCTGAA GGGGATAAGC
CGCGCAGCGC TGAAGCTATT TTGATGACGA ATCACTTTAA AACCAGCACG CTGTTTTGTG
CCTCCATGCA GATGGCCTCG ATTGTTGCGA ATGCCTCCAG CGAAGCGCGT GATTGCCTGC

ATCGTTTTTC ACTTGATCTT GGTCAGGCAT TTCAACTGCT GGACGATTTG ACCGATGGCA
TGACCGACAC CGGTAAGGAT AGCAATCAGG ACGCCGGTAA ATCGACGCTG GTCAATCTGT
TAGGCCCGAG GGCGGTTGAA GAACGTCTGA GACAACATCT TCAGCTTGCC AGTGAGCATC
TCTCTGCGGC CTGCCAACAC GGGCACGCCA CTCAACATTT TATTCAGGCC TGGTTTGACA
AAAAACTCGC TGCCGTCAGT TAAGCTTATG TGCACCGGTC AGCCTGTCTT AAGTGGGAGC
GGCTATGCAA CCGCATTATG ATCTGATTCT CGTGGGGGCT GGACTCGCGA ATGGCCTTAT
CGCCCTGCGT CTTCAGCAGC AGCAACCTGA TATGCGTATT TTGCTTATCG ACGCCGCACC
CCAGGCGGGC GGGAATCATA CGTGGTCATT TCACCACGAT GATTTGACTG AGAGCCAACA
TCGTTGGATA GCTCCGCTGG TGGTTCATCA CTGGCCCGAC TATCAGGTAC GCTTTCCCAC
ACGCCGTCGT AAGCTGAACA GCGGCTACTT TTGTATTACT TCTCAGCGTT TCGCTGAGGT
TTTACAGCGA CAGTTTGGCC CGCACTTGTG GATGGATACC GCGGTCGCAG AGGTTAATGC
GGAATCTGTT CGGTTGAAAA AGGGTCAGGT TATCGGTGCC CGCGCGGTGA TTGACGGGCG
GGGTTATGCG GCAAATTCAG CACTGAGCGT GGGCTTCCAG GCGTTTATTG GCCAGGAATG
GCGATTGAGC CACCCGCATG GTTTATCGTC TCCCATTATC ATGGATGCCA CGGTCGATCA
GCAAAATGGT TATCGCTTCG TGTACAGCCT GCCGCTCTCG CCGACCAGAT TGTTAATTGA
AGACACGCAC TATATTGATA ATGCGACATT AGATCCTGAA TGCGCGCGGC AAAATATTTG
CGACTATGCC GCGCAACAGG GTTGGCAGCT TCAGACACTG CTGCGAGAAG AACAGGGCGC
CTTACCCATT ACTCTGTCGG GCAATGCCGA CGCATTCTGG CAGCAGCGCC CCCTGGCCTG
TAGTGGATTA CGTGCCGGTC TGTTCCATCC TACCACCGGC TATTCACTGC CGCTGGCGGT
TGCCGTGGCC GACCGCCTGA GTGCACTTGA TGTCTTTACG TCGGCCTCAA TTCACCATGC
CATTACGCAT TTTGCCCGCG AGCGCTGGCA GCAGCAGGGC TTTTTCCGCA TGCTGAATCG
CATGCTGTTT TTAGCCGGAC CCGCCGATTC ACGCTGGCGG GTTATGCAGC GTTTTTATGG
TTTACCTGAA GATTTAATTG CCCGTTTTTA TGCGGGAAAA CTCACGCTGA CCGATCGGCT
ACGTATTCTG AGCGGCAAGC CGCCTGTTCC GGTATTAGCA GCATTGCAAG CCATTATGAC
GACTCATCGT TAAAGAGCGA CTACATGAAA CCAACTACGG TAATTGGTGC AGGCTTCGGT
GGCCTGGCAC TGGCAATTCG TCTACAAGCT GCGGGGATCC CCGTCTTACT GCTTGAACAA
CGTGATAAAC CCGGCGGTCG GGCTTATGTC TACGAGGATC AGGGGTTTAC CTTTGATGCA
GGCCCGACGG TTATCACCGA TCCCAGTGCC ATTGAAGAAC TGTTTGCACT GGCAGGAAAA
CAGTTAAAAG AGTATGTCGA ACTGCTGCCG GTTACGCCGT TTTACCGCCT GTGTTGGGAG
TCAGGGAAGG TCTTTAATTA CGATAACGAT CAAACCCGGC TCGAAGCGCA GATTCAGCAG
TTTAATCCCC GCGATGTCGA AGGTTATCGT CAGTTTCTGG ACTATTCACG CGCGGTGTTT
AAAGAAGGCT ATCTAAAGCT CGGTACTGTC CCTTTTTTAT CGTTCAGAGA CATGCTTCGC
GCCGCACCTC AACTGGCGAA ACTGCAGGCA TGGAGAAGCG TTTACAGTAA GGTTGCCAGT
TACATCGAAG ATGAACATCT GCGCCAGGCG TTTTCTTTCC ACTCGCTGTT GGTGGGCGGC
AATCCCTTCG CCACCTCATC CATTTATACG TTGATACACG CGCTGGAGCG TGAGTGGGGC
GTCTGGTTTC CGCGTGGCGG CACCGGCGCA TTAGTTCAGG GGATGATAAA GCTGTTTCAG
GATCTGGGTG GCGAAGTCGT GTTAAACGCC AGAGTCAGCC ATATGGAAAC GACAGGAAAC
AAGATTGAAG CCGTGCATTT AGAGGACGGT CGCAGGTTCC TGACGCAAGC CGTCGCGTCA
AATGCAGATG TGGTTCATAC CTATCGCGAC CTGTTAAGCC AGCACCCTGC CGCGGTTAAG
CAGTCCAACA AACTGCAGAC TAAGCGCATG AGTAACTCTC TGTTTGTGCT CTATTTTGGT
TTGAATCACC ATCATGATCA GCTCGCGCAT CACACGGTTT GTTTCGGCCC GCGTTACCGC
GAGCTGATTG ACGAAATTTT TAATCATGAT GGCCTCGCAG AGGACTTCTC ACTTTATCTG
CACGCGCCCT GTGTCACGGA TTCGTCACTG GCGCCTGAAG GTTGCGGCAG TTACTATGTG
TTGGCGCCGG TGCCGCATTT AGGCACCGCG AACCTCGACT GGACGGTTGA GGGGCCAAAA
CTACGCGACC GTATTTTTGC GTACCTTGAG CAGCATTACA TGCCTGGCTT ACGGAGTCAG
CTGGTCACGC ACCGGATGTT TACGCCGTTT GATTTTCGCG ACCAGCTTAA TGCCTATCAT
GGCTCAGCCT TTTCTGTGGA GCCCGTTCTT ACCCAGAGCG CCTGGTTTCG GCCGCATAAC
CGCGATAAAA CCATTACTAA TCTCTACCTG GTCGGCGCAG GCACGCATCC CGGCGCAGGC
ATTCCTGGCG TCATCGGCTC GGCAAAAGCG ACAGCAGGTT TGATGCTGGA GGATCTGATT
TGAATAATCC GTCGTTACTC AATCATGCGG TCGAAACGAT GGCAGTTGGC TCGAAAAGTT
TTGCGACAGC CTCAAAGTTA TTTGATGCAA AAACCCGGCG CAGCGTACTG ATGCTCTACG
CCTGGTGCCG CCATTGTGAC GATGTTATTG ACGATCAGAC GCTGGGCTTT CAGGCCCGGC
AGCCTGCCTT ACAAACGCCC GAACAACGTC TGATGCAACT TGAGATGAAA ACGCGCCAGG
CCTATGCAGG ATCGCAGATG CACGAACCGG CGTTTGCGGC TTTTCAGGAA GTGGCTATGG
CTCATGATAT CGCCCCGGCT TACGCGTTTG ATCATCTGGA AGGCTTCGCC ATGGATGTAC

GCGAAGCGCA ATACAGCCAA CTGGATGATA CGCTGCGCTA TTGCTATCAC GTTGCAGGCG
TTGTCGGCTT GATGATGGCG CAAATCATGG GCGTGCGGGA TAACGCCACG CTGGACCGCG
CCTGTGACCT TGGGCTGGCA TTTCAGTTGA CCAATATTGC TCGCGATATT GTGGACGATG
CGCATGCGGG CCGCTGTTAT CTGCCGGCAA GCTGGCTGGA GCATGAAGGT CTGAACAAAG
AGAATTATGC GGCACCTGAA AACCGTCAGG CGCTGAGCCG TATCGCCCGT CGTTTGGTGC
AGGAAGCAGA ACCTTACTAT TTGTCTGCCA CAGCCGGCCT GGCAGGGTTG CCCCTGCGTT
CCGCCTGGGC AATCGCTACG GCGAAGCAGG TTTACCGGAA AATAGGTGTC AAAGTTGAAC
AGGCCGGTCA GCAAGCCTGG GATCAGCGGC AGTCAACGAC CACGCCCGAA AAATTAACGC
TGCTGCTGGC CGCCTCTGGT CAGGCCCTTA CTTCCCGGAT GCGGGCTCAT CCTCCCCGCC
CTGCGCATCT CTGGCAGCGC CCGCTCTAGC GCCATGTCGA CCTCGAGGGG GGGCCCGGTA
CCCAATTCGC CCTATAGTGA GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT
CGTGACTGGG AAAACCCTGG CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCTTTC
GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC CTTCCCAACA GTTGCGCAGC
CTGAATGGCG AATGGAAATT GTAAGCGTTA ATATTTTGTT AAAATTCGCG TTAAATTTTT
GTTAAATCAG CTCATTTTTT AACCAATAGG CCGAAATCGG CAAAATCCCT TATAAATCAA
AAGAATAGAC CGAGATAGGG TTGAGTGTTG TTCCAGTTTG GAACAAGAGT CCACTATTAA
AGAACGTGGA CTCCAACGTC AAAGGGCGAA AAACCGTCTA TCAGGGCGAT GGCCCACTAC
GTGAACCATC ACCCTAATCA AGTTTTTTGG GGTCGAGGTG CCGTAAAGCA CTAAATCGGA
ACCCTAAAGG GAGCCCCCGA TTTAGAGCTT GACGGGGAAA GCCGGCGAAC GTGGCGAGAA
AGGAAGGGAA GAAAGCGAAA GGAGCGGGCG CTAGGGCGCT GGCAAGTGTA GCGGTCACGC
TGCGCGTAAC CACCACACCC GCCGCGCTTA ATGCGCCGCT ACAGGGCGCG TCAG

//

Annex F

DNA sequence of the vector construct pACYCcaro14

ID PACYCCARO14 PRELIMINARY; DNA; 8547 BP.
SQ SEQUENCE 8547 BP; 1884 A; 2296 C; 2279 G; 2088 T; 0 OTHER;

GAATTCCGGA TGAGCATTCA TCAGGCGGGC AAGAATGTGA ATAAAGGCCG GATAAAACTT
GTGCTTATTT TTCTTTACGG TCTTTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT
ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA
TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA
AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT GGTGAAAGTT
GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC
GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG TCACAGGTAT
TTATTCGGCG CAAAGTGCGT CGGGTGATGC TGCCAACTTA CTGATTTAGT GTATGATGGT
GTTTTTGAGG TGCTCCAGTG GCTTCTGTTT CTATCAGCTG TCCCTCCTGT TCAGCTACTG
ACGGGGTGGT GCGTAACGGC AAAAGCACCG CCGGACATCA GCGCTAGCGG AGTGTATACT
GGCTTACTAT GTTGGCACTG ATGAGGGTGT CAGTGAAGTG CTTCATGTGG CAGGAGAAAA
AAGGCTGCAC CGGTGCGTCA GCAGAATATG TGATACAGGA TATATTCCGC TTCCTCGCTC
ACTGACTCGC TACGCTCGGT CGTTCGACTG CGGCGAGCGG AAATGGCTTA CGAACGGGGC
GGAGATTTCC TGGAAGATGC CAGGAAGATA CTTAACAGGG AAGTGAGAGG GCCGCGGCAA
AGCCGTTTTT CCATAGGCTC CGCCCCCCTG ACAAGCATCA CGAAATCTGA CGCTCAAATC
AGTGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC GTTTCCCCCT GGCGGCTCCC
TCGTGCGCTC TCCTGTTCCT GCCTTTCGGT TTACCGGTGT CATTCCGCTG TTATGGCCGC
GTTTGTCTCA TTCCACGCCT GACACTCAGT TCCGGGTAGG CAGTTCGCTC CAAGCTGGAC
TGTATGCACG AACCCCCCGT TCAGTCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
GAGTCCAACC CGGAAAGACA TGCAAAAGCA CCACTGGCAG CAGCCACTGG TAATTGATTT
AGAGGAGTTA GTCTTGAAGT CATGCGCCGG TTAAGGCTAA ACTGAAAGGA CAAGTTTTGG
TGACTGCGCT CCTCCAAGCC AGTTACCTCG GTTCAAAGAG TTGGTAGCTC AGAGAACCTT
CGAAAAACCG CCCTGCAAGG CGGTTTTTTC GTTTTCAGAG CAAGAGATTA CGCGCAGACC
AAAACGATCT CAAGAAGATC ATCTTATTAA TCAGATAAAA TATTTCTAGA TTTCAGTGCA
ATTTATCTCT TCAAATGTAG CACCTGAAGT CAGCCCCATA CGATATAAGT TGTAATTCTC
ATGTTTGACA GCTTATCATC GATAAGCTTT AATGCGGTAG TTTATCACAG TTAAATTGCT
AACGCAGTCA GGCACCGTGT ATGAAATCTA ACAATGCGCT CATCGTCATC CTCGGCACCG
TCACCCTGGA TGCTGTAGGC ATAGGCTTGG TTATGCCGGT ACTGCCGGGC CTCTTGCGGG
ATATCGTCCA TTCCGACAGC ATCGCCAGTC ACTATGGCGT GCTGCTAGCG CTATATGCGT
TGATGCAATT TCTATGCGCA CCCGTTCTCG GAGCACTGTC CGACCGCTTT GGCCGCCGCC
CAGTCCTGCT CGCTTCGCTA CTTGGAGCCA CTATCGACTA CGCGATCATG GCGACCACAC
CCGTCCTGTG GATCCCCCGG GCTGCAGGAA TTGCCGTAAA TGTATCCGTT TATAAGGACA
GCCCGAATGA CGGTCTGCGC AAAAAAACAC GTTCATCTCA CTCGCGATGC TGCGGAGCAG
TTACTGGCTG ATATTGATCG ACGCCTTGAT CAGTTATTGC CCGTGGAGGG AGAACGGGAT
GTTGTGGGTG CCGCGATGCG TGAAGGTGCG CTGGCACCGG GAAAACGTAT TCGCCCCATG
TTGCTGTTGC TGACCGCCCG CGATCTGGGT TGCGCTGTCA GCCATGACGG ATTACTGGAT
TTGGCCTGTG CGGTGGAAAT GGTCCACGCG GCTTCGCTGA TCCTTGACGA TATGCCCTGC
ATGGACGATG CGAAGCTGCG GCGCGGACGC CCTACCATTC ATTCTCATTA CGGAGAGCAT
GTGGCAATAC TGGCGGCGGT TGCCTTGCTG AGTAAAGCCT TTGGCGTAAT TGCCGATGCA
GATGGCCTCA CGCCGCTGGC AAAAAATCGG GCGGTTTCTG AACTGTCAAA CGCCATCGGC
ATGCAAGGAT TGGTTCAGGG TCAGTTCAAG GATCTGTCTG AAGGGGATAA GCCGCGCAGC
GCTGAAGCTA TTTTGATGAC GAATCACTTT AAAACCAGCA CGCTGTTTTG TGCCTCCATG
CAGATGGCCT CGATTGTTGC GAATGCCTCC AGCGAAGCGC GTGATTGCCT GCATCGTTTT
TCACTTGATC TTGGTCAGGC ATTTCAACTG CTGGACGATT TGACCGATGG CATGACCGAC
ACCGGTAAGG ATAGCAATCA GGACGCCGGT AAATCGACGC TGGTCAATCT GTTAGGCCCG
AGGGCGGTTG AAGAACGTCT GAGACAACAT CTTCAGCTTG CCAGTGAGCA TCTCTCTGCG
GCCTGCCAAC ACGGGCACGC CACTCAACAT TTTATTCAGG CCTGGTTTGA CAAAAAACTC
GCTGCCGTCA GTTAAGCTTA TGTGCACCGG TCAGCCTGTC TTAAGTGGGA GCGGCTATGC
AACCGCATTA TGATCTGATT CTCGTGGGGG CTGGACTCGC GAATGGCCTT ATCGCCCTGC

GTCTTCAGCA GCAGCAACCT GATATGCGTA TTTTGCTTAT CGACGCCGCA CCCCAGGCGG
GCGGGAATCA TACGTGGTCA TTTCACCACG ATGATTTGAC TGAGAGCCAA CATCGTTGGA
TAGCTCCGCT GGTGGTTCAT CACTGGCCCG ACTATCAGGT ACGCTTTCCC ACACGCCGTC
GTAAGCTGAA CAGCGGCTAC TTTTGTATTA CTTCTCAGCG TTTCGCTGAG GTTTTACAGC
GACAGTTTGG CCCGCACTTG TGGATGGATA CCGCGGTCGC AGAGGTTAAT GCGGAATCTG
TTCGGTTGAA AAAGGGTCAG GTTATCGGTG CCCGCGCGGT GATTGACGGG CGGGGTTATG
CGGCAAATTC AGCACTGAGC GTGGGCTTCC AGGCGTTTAT TGGCCAGGAA TGGCGATTGA
GCCACCCGCA TGGTTTATCG TCTCCCATTA TCATGGATGC CACGGTCGAT CAGCAAAATG
GTTATCGCTT CGTGTACAGC CTGCCGCTCT CGCCGACCAG ATTGTTAATT GAAGACACGC
ACTATATTGA TAATGCGACA TTAGATCCTG AATGCGCGCG GCAAAATATT TGCGACTATG
CCGCGCAACA GGGTTGGCAG CTTCAGACAC TGCTGCGAGA AGAACAGGGC GCCTTACCCA
TTACTCTGTC GGGCAATGCC GACGCATTCT GGCAGCAGCG CCCCCTGGCC TGTAGTGGAT
TACGTGCCGG TCTGTTCCAT CCTACCACCG GCTATTCACT GCCGCTGGCG GTTGCCGTGG
CCGACCGCCT GAGTGCACTT GATGTCTTTA CGTCGGCCTC AATTCACCAT GCCATTACGC
ATTTTGCCCG CGAGCGCTGG CAGCAGCAGG GCTTTTTCCG CATGCTGAAT CGCATGCTGT
TTTTAGCCGG ACCCGCCGAT TCACGCTGGC GGGTTATGCA GCGTTTTTAT GGTTTACCTG
AAGATTTAAT TGCCCGTTTT TATGCGGGAA AACTCACGCT GACCGATCGG CTACGTATTC
TGAGCGGCAA GCCGCCTGTT CCGGTATTAG CAGCATTGCA AGCCATTATG ACGACTCATC
GTTAAAGAGC GACTACATGA AACCAACTAC GGTAATTGGT GCAGGCTTCG GTGGCCTGGC
ACTGGCAATT CGTCTACAAG CTGCGGGGAT CCCCGTCTTA CTGCTTGAAC AACGTGATAA
ACCCGGCGGT CGGGCTTATG TCTACGAGGA TCAGGGGTTT ACCTTTGATG CAGGCCCGAC
GGTTATCACC GATCCCAGTG CCATTGAAGA ACTGTTTGCA CTGGCAGGAA AACAGTTAAA
AGAGTATGTC GAACTGCTGC CGGTTACGCC GTTTTACCGC CTGTGTTGGG AGTCAGGGAA
GGTCTTTAAT TACGATAACG ATCAAACCCG GCTCGAAGCG CAGATTCAGC AGTTTAATCC
CCGCGATGTC GAAGGTTATC GTCAGTTTCT GGACTATTCA CGCGCGGTGT TTAAAGAAGG
CTATCTAAAG CTCGGTACTG TCCCTTTTTT ATCGTTCAGA GACATGCTTC GCGCCGCACC
TCAACTGGCG AAACTGCAGG CATGGAGAAG CGTTTACAGT AAGGTTGCCA GTTACATCGA
AGATGAACAT CTGCGCCAGG CGTTTTCTTT CCACTCGCTG TTGGTGGGCG GCAATCCCTT
CGCCACCTCA TCCATTTATA CGTTGATACA CGCGCTGGAG CGTGAGTGGG GCGTCTGGTT
TCCGCGTGGC GGCACCGGCG CATTAGTTCA GGGGATGATA AAGCTGTTTC AGGATCTGGG
TGGCGAAGTC GTGTTAAACG CCAGAGTCAG CCATATGGAA ACGACAGGAA ACAAGATTGA
AGCCGTGCAT TTAGAGGACG GTCGCAGGTT CCTGACGCAA GCCGTCGCGT CAAATGCAGA
TGTGGTTCAT ACCTATCGCG ACCTGTTAAG CCAGCACCCT GCCGCGGTTA AGCAGTCCAA
CAAACTGCAG ACTAAGCGCA TGAGTAACTC TCTGTTTGTG CTCTATTTTG GTTTGAATCA
CCATCATGAT CAGCTCGCGC ATCACACGGT TTGTTTCGGC CCGCGTTACC GCGAGCTGAT
TGACGAAATT TTTAATCATG ATGGCCTCGC AGAGGACTTC TCACTTTATC TGCACGCGCC
CTGTGTCACG GATTCGTCAC TGGCGCCTGA AGGTTGCGGC AGTTACTATG TGTTGGCGCC
GGTGCCGCAT TTAGGCACCG CGAACCTCGA CTGGACGGTT GAGGGGCCAA AACTACGCGA
CCGTATTTTT GCGTACCTTG AGCAGCATTA CATGCCTGGC TTACGGAGTC AGCTGGTCAC
GCACCGGATG TTTACGCCGT TTGATTTTCG CGACCAGCTT AATGCCTATC ATGGCTCAGC
CTTTTCTGTG GAGCCCGTTC TTACCCAGAG CGCCTGGTTT CGGCCGCATA ACCGCGATAA
AACCATTACT AATCTCTACC TGGTCGGCGC AGGCACGCAT CCCGGCGCAG GCATTCCTGG
CGTCATCGGC TCGGCAAAAG CGACAGCAGG TTTGATGCTG GAGGATCTGA TTTGAATAAT
CCGTCGTTAC TCAATCATGC GGTCGAAACG ATGGCAGTTG GCTCGAAAAG TTTTGCGACA
GCCTCAAAGT TATTTGATGC AAAAACCCGG CGCAGCGTAC TGATGCTCTA CGCCTGGTGC
CGCCATTGTG ACGATGTTAT TGACGATCAG ACGCTGGGCT TTCAGGCCCG GCAGCCTGCC
TTACAAACGC CCGAACAACG TCTGATGCAA CTTGAGATGA AAACGCGCCA GGCCTATGCA
GGATCGCAGA TGCACGAACC GGCGTTTGCG GCTTTTCAGG AAGTGGCTAT GGCTCATGAT
ATCGCCCCGG CTTACGCGTT TGATCATCTG GAAGGCTTCG CCATGGATGT ACGCGAAGCG
CAATACAGCC AACTGGATGA TACGCTGCGC TATTGCTATC ACGTTGCAGG CGTTGTCGGC
TTGATGATGG CGCAAATCAT GGGCGTGCGG GATAACGCCA CGCTGGACCG CGCCTGTGAC
CTTGGGCTGG CATTTCAGTT GACCAATATT GCTCGCGATA TTGTGGACGA TGCGCATGCG
GGCCGCTGTT ATCTGCCGGC AAGCTGGCTG GAGCATGAAG GTCTGAACAA AGAGAATTAT
GCGGCACCTG AAAACCGTCA GGCGCTGAGC CGTATCGCCC GTCGTTTGGT GCAGGAAGCA
GAACCTTACT ATTTGTCTGC CACAGCCGGC CTGGCAGGGT TGCCCCTGCG TTCCGCCTGG

GCAATCGCTA CGGCGAAGCA GGTTTACCGG AAAATAGGTG TCAAAGTTGA ACAGGCCGGT
CAGCAAGCCT GGGATCAGCG GCAGTCAACG ACCACGCCCG AAAAATTAAC GCTGCTGCTG
GCCGCCTCTG GTCAGGCCCT TACTTCCCGG ATGCGGGCTC ATCCTCCCCG CCCTGCGCAT
CTCTGGCAGC GCCCGCTCTA GCGCCATGTC GACCGATGCC CTTGAGAGCC TTCAACCCAG
TCAGCTCCTT CCGGTGGGCG CGGGGCATGA CTATCGTCGC CGCACTTATG ACTGTCTTCT
TTATCATGCA ACTCGTAGGA CAGGTGCCGG CAGCGCTCTG GGTCATTTTC GGCGAGGACC
GCTTTCGCTG GAGCGCGACG ATGATCGGCC TGTCGCTTGC GGTATTCGGA ATCTTGCACG
CCCTCGCTCA AGCCTTCGTC ACTGGTCCCG CCACCAAACG TTTCGGCGAG AAGCAGGCCA
TTATCGCCGG CATGGCGGCC GACGCGCTGG GCTACGTCTT GCTGGCGTTC GCGACGCGAG
GCTGGATGGC CTTCCCCATT ATGATTCTTC TCGCTTCCGG CGGCATCGGG ATGCCCGCGT
TGCAGGCCAT GCTGTCCAGG CAGGTAGATG ACGACCATCA GGGACAGCTT CAAGGATCGC
TCGCGGCTCT TACCAGCCTA ACTTCGATCA CTGGACCGCT GATCGTCACG GCGATTTATG
CCGCCTCGGC GAGCACATGG AACGGGTTGG CATGGATTGT AGGCGCCGCC CTATACCTTG
TCTGCCTCCC CGCGTTGCGT CGCGGTGCAT GGAGCCGGGC CACCTCGACC TGAATGGAAG
CCGGCGGCAC CTCGCTAACG GATTCACCAC TCCAAGAATT GGAGCCAATC AATTCTTGCG
GAGAACTGTG AATGCGCAAA CCAACCCTTG GCAGAACATA TCCATCGCGT CCGCCATCTC
CAGCAGCCGC ACGCGGCGCA TCTCGGGCAG CGTTGGGTCC TGGCCACGGG TGCGCATGAT
CGTGCTCCTG TCGTTGAGGA CCCGGCTAGG CTGGCGGGGT TGCCTTACTG GTTAGCAGAA
TGAATCACCG ATACGCGAGC GAACGTGAAG CGACTGCTGC TGCAAAACGT CTGCGACCTG
AGCAACAACA TGAATGGTCT TCGGTTTCCG TGTTTCGTAA AGTCTGGAAA CGCGGAAGTC
CCCTACGTGC TGCTGAAGTT GCCCGCAACA GAGAGTGGAA CCAACCGGTG ATACCACGAT
ACTATGACTG AGAGTCAACG CCATGAGCGG CCTCATTTCT TATTCTGAGT TACAACAGTC
CGCACCGCTG TCCGGTAGCT CCTTCCGGTG GGCGCGGGGC ATGACTATCG TCGCCGCACT
TATGACTGTC TTCTTTATCA TGCAACTCGT AGGACAGGTG CCGGCAGCGC CCAACAGTCC
CCCGGCCACG GGGCCTGCCA CCATACCCAC GCCGAAACAA GCGCCCTGCA CCATTATGTT
CCGGATCTGC ATCGCAGGAT GCTGCTGGCT ACCCTGTGGA ACACCTACAT CTGTATTAAC
GAAGCGCTAA CCGTTTTTAT CAGGCTCTGG GAGGCAGAAT AAATGATCAT ATCGTCAATT
ATTACCTCCA CGGGGAGAGC CTGAGCAAAC TGGCCTCAGG CATTTGAGAA GCACACGGTC
ACACTGCTTC CGGTAGTCAA TAAACCGGTA AACCAGCAAT AGACATAAGC GGCTATTTAA
CGACCCTGCC CTGAACCGAC GACCGGGTCG AATTTGCTTT CGAATTTCTG CCATTCATCC
GCTTATTATC ACTTATTCAG GCGTAGCACC AGGCGTTTAA GGGCACCAAT AACTGCCTTA
AAAAAATTAC GCCCCGCCCT GCCACTCATC GCAGTACTGT TGTAATTCAT TAAGCATTCT
GCCGACATGG AAGCCATCAC AGACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC
CTTGTCGCCT TGCGTATAAT ATTTGCCCAT GGTGAAAACG GGGGCGAAGA AGTTGTCCAT
ATTGGCCACG TTTAAATCAA AACTGGTGAA ACTCACCCAG GGATTGGCTG AGACGAAAAA
CATATTCTCA ATAAACCCTT TAGGGAAATA GGCCAGGTTT TCACCGTAAC ACGCCACATC
TTGCGAATAT ATGTGTAGAA ACTGCCGGAA ATCGTCGTGG TATTCACTCC AGAGCGATGA
AAACGTTTCA GTTTGCTCAT GGAAAACGGT GTAACAAGGG TGAACACTAT CCCATATCAC
CAGCTCACCG TCTTTCATTG CCATACG

//

Annex G

DNA and corresponding amino acid sequence of the *ispG* (formely *gcpE*) gene of *Escherichia coli*

```
        10         20         30         40         50         60
         |          |          |          |          |          |
ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG
 M  H  N  Q  A  P  I  Q  R  R  K  S  T  R  I  Y  V  G  N  V

        70         80         90        100        110        120
         |          |          |          |          |          |
CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC
 P  I  G  D  G  A  P  I  A  V  Q  S  M  T  N  T  R  T  T  D

       130        140        150        160        170        180
         |          |          |          |          |          |
GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT
 V  E  A  T  V  N  Q  I  K  A  L  E  R  V  G  A  D  I  V  R

       190        200        210        220        230        240
         |          |          |          |          |          |
GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC
 V  S  V  P  T  M  D  A  A  E  A  F  K  L  I  K  Q  Q  V  N

       250        260        270        280        290        300
         |          |          |          |          |          |
GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC
 V  P  L  V  A  D  I  H  F  D  Y  R  I  A  L  K  V  A  E  Y

       310        320        330        340        350        360
         |          |          |          |          |          |
GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG
 G  V  D  C  L  R  I  N  P  G  N  I  G  N  E  E  R  I  R  M

       370        380        390        400        410        420
         |          |          |          |          |          |
GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG
 V  V  D  C  A  R  D  K  N  I  P  I  R  I  G  V  N  A  G  S

       430        440        450        460        470        480
         |          |          |          |          |          |
CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT
 L  E  K  D  L  Q  E  K  Y  G  E  P  T  P  Q  A  L  L  E  S

       490        500        510        520        530        540
         |          |          |          |          |          |
GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG
 A  M  R  H  V  D  H  L  D  R  L  N  F  D  Q  F  K  V  S  V
```

```
       550        560        570        580        590        600
        |          |          |          |          |          |
AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT
 K  A  S  D  V  F  L  A  V  E  S  Y  R  L  L  A  K  Q  I  D

       610        620        630        640        650        660
        |          |          |          |          |          |
CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC
 Q  P  L  H  L  G  I  T  E  A  G  G  A  R  S  G  A  V  K  S

       670        680        690        700        710        720
        |          |          |          |          |          |
GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG
 A  I  G  L  G  L  L  L  S  E  G  I  G  D  T  L  R  V  S  L

       730        740        750        760        770        780
        |          |          |          |          |          |
GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC

 A  A  D  P  V  E  E  I  K  V  G  F  D  I  L  K  S  L  R  I

       790        800        810        820        830        840
        |          |          |          |          |          |
CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT
 R  S  R  G  I  N  F  I  A  C  P  T  C  S  R  Q  E  F  D  V

       850        860        870        880        890        900
        |          |          |          |          |          |
ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT
 I  G  T  V  N  A  L  E  Q  R  L  E  D  I  I  T  P  M  D  V

       910        920        930        940        950        960
        |          |          |          |          |          |
TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC
 S  I  I  G  C  V  V  N  G  P  G  E  A  L  V  S  T  L  G  V

       970        980        990       1000       1010       1020
        |          |          |          |          |          |
ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC
 T  G  G  N  K  K  S  G  L  Y  E  D  G  V  R  K  D  R  L  D

      1030       1040       1050       1060       1070       1080
        |          |          |          |          |          |
AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC
 N  N  D  M  I  D  Q  L  E  A  R  I  R  A  K  A  S  Q  L  D

      1090       1100       1110
        |          |          |
GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAATAA
 E  A  R  R  I  D  V  Q  Q  V  E  K  -
```

Annex H

DNA sequence of the vector construct pBScyclogcpE

ID PBSCYCLOG PRELIMINARY; DNA; 8823 BP.

SQ SEQUENCE 8823 BP; 2123 A; 2169 C; 2468 G; 2063 T; 0 OTHER;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGGAGGA
GAAATTAACC ATGCATAACC AGGCTCCAAT TCAACGTAGA AAATCAACAC GTATTTACGT
TGGGAATGTG CCGATTGGCG ATGGTGCTCC CATCGCCGTA CAGTCCATGA CCAATACGCG
TACGACAGAC GTCGAAGCAA CGGTCAATCA AATCAAGGCG CTGGAACGCG TTGGCGCTGA
TATCGTCCGT GTATCCGTAC CGACGATGGA CGCGGCAGAA GCGTTCAAAC TCATCAAACA
GCAGGTTAAC GTGCCGCTGG TGGCTGACAT CCACTTCGAC TATCGCATTG CGCTGAAAGT
AGCGGAATAC GGCGTCGATT GTCTGCGTAT TAACCCTGGC AATATCGGTA ATGAAGAGCG
TATTCGCATG GTGGTTGACT GTGCGCGCGA TAAAAACATT CCGATCCGTA TTGGCGTTAA
CGCCGGATCG CTGGAAAAAG ATCTGCAAGA AAAGTATGGC GAACCGACGC CGCAGGCGTT

GCTGGAATCT GCCATGCGTC ATGTTGATCA TCTCGATCGC CTGAACTTCG ATCAGTTCAA
AGTCAGCGTG AAAGCGTCTG ACGTCTTCCT CGCTGTTGAG TCTTATCGTT TGCTGGCAAA
ACAGATCGAT CAGCCGTTGC ATCTGGGGAT CACCGAAGCC GGTGGTGCGC GCAGCGGGGC
AGTAAAATCC GCCATTGGTT TAGGTCTGCT GCTGTCTGAA GGCATCGGCG ACACGCTGCG
CGTATCGCTG GCGGCCGATC CGGTCGAAGA GATCAAAGTC GGTTTCGATA TTTTGAAATC
GCTGCGTATC CGTTCGCGAG GGATCAACTT CATCGCCTGC CCGACCTGTT CGCGTCAGGA
ATTTGATGTT ATCGGTACGG TTAACGCGCT GGAGCAACGC CTGGAAGATA TCATCACTCC
GATGGACGTT TCGATTATCG GCTGCGTGGT GAATGGCCCA GGTGAGGCGC TGGTTTCTAC
ACTCGGCGTC ACCGGCGGCA ACAAGAAAAG CGGCCTCTAT GAAGATGGCG TGCGCAAAGA
CCGTCTGGAC AACAACGATA TGATCGACCA GCTGGAAGCA CGCATTCGTG CGAAAGCCAG
TCAGCTGGAC GAAGCGCGTC GAATTGACGT TCAGCAGGTT GAAAAATAAG CGGCCGCTCT
AGAACTAGTG GATCCCCCGG GCTGCAGGAA TTCGAGGAGA AATTAACCAT GTATATCGGG
ATAGATCTTG GCACCTCGGG CGTAAAAGTT ATTTTGCTCA ACGAGCAGGG TGAGGTGGTT
GCTGCGCAAA CGGAAAAGCT GACCGTTTCG CGCCCGCATC CACTCTGGTC GGAACAAGAC
CCGGAACAGT GGTGGCAGGC AACTGATCGC GCAATGAAAG CTCTGGGCGA TCAGCATTCT
CTGCAGGACG TTAAAGCATT GGGTATTGCC GGCCAGATGC ACGGAGCAAC CTTGCTGGAT
GCTCAGCAAC GGGTGTTACG CCCTGCCATT TTGTGGAACG ACGGGCGCTG TGCGCAAGAG
TGCACTTTGC TGGAAGCGCG AGTTCCGCAA TCGCGGGTGA TTACCGGCAA CCTGATGATG
CCCGGATTTA CTGCGCCTAA ATTGCTATGG GTTCAGCGGC ATGAGCCGGA GATATTCCGT
CAAATCGACA AAGTATTATT ACCGAAAGAT TACTTGCGTC TGCGTATGAC GGGGGAGTTT
GCCAGCGATA TGTCTGACGC AGCTGGCACC ATGTGGCTGG ATGTCGCAAA GCGTGACTGG
AGTGACGTCA TGCTGCAGGC TTGCGACTTA TCTCGTGACC AGATGCCCGC ATTATACGAA
GGCAGCGAAA TTACTGGTGC TTTGTTACCT GAAGTTGCGA AAGCGTGGGG TATGGCGACG
GTGCCAGTTG TCGCAGGCGG TGGCGACAAT GCAGCTGGTG CAGTTGGTGT GGGAATGGTT
GATGCTAATC AGGCAATGTT ATCGCTGGGG ACGTCGGGGG TCTATTTTGC TGTCAGCGAA
GGGTTCTTAA GCAAGCCAGA AAGCGCCGTA CATAGCTTTT GCCATGCGCT ACCGCAACGT
TGGCATTTAA TGTCTGTGAT GCTGAGTGCA GCGTCGTGTC TGGATTGGGC CGCGAAATTA
ACCGGCCTGA GCAATGTCCC AGCTTTAATC GCTGCAGCTC AACAGGCTGA TGAAAGTGCC
GAGCCAGTTT GGTTTCTGCC TTATCTTTCC GGCGAGCGTA CGCCACACAA TAATCCCCAG
GCGAAGGGGG TTTTCTTTGG TTTGACTCAT CAACATGGCC CCAATGAACT GGCGCGAGCA
GTGCTGGAAG GCGTGGGTTA TGCGCTGGCA GATGGCATGG ATGTCGTGCA TGCCTGCGGT
ATTAAACCGC AAAGTGTTAC GTTGATTGGG GGCGGGGCGC GTAGTGAGTA CTGGCGTCAG
ATGCTGGCGG ATATCAGCGG TCAGCAGCTC GATTACCGTA CGGGGGGGGA TGTGGGGCCA
GCACTGGGCG CAGCAAGGCT GGCGCAGATC GCGGCGAATC CAGAGAAATC GCTCATTGAA
TTGTTGCCGC AACTACCGTT AGAACAGTCG CATCTACCAG ATGCGCAGCG TTATGCCGCT
TATCAGCCAC GACGAGAAAC GTTCCGTCGC CTCTATCAGC AACTTCTGCC ATTAATGGCG
TAAAAGCTTG AGGAGAAATT AACCATGAAG CAACTCACCA TTCTGGGCTC GACCGGCTCG
ATTGGTTGCA GCACGCTGGA CGTGGTGCGC CATAATCCCG AACACTTCCG CGTAGTTGCG
CTGGTGGCAG GCAAAAATGT CACTCGCATG GTAGAACAGT GCCTGGAATT CTCTCCCCGC
TATGCCGTAA TGGACGATGA AGCGAGTGCG AAACTTCTTA AAACGATGCT ACAGCAACAG
GGTAGCCGCA CCGAAGTCTT AAGTGGGCAA CAAGCCGCTT GCGATATGGC AGCGCTTGAG
GATGTTGATC AGGTGATGGC AGCCATTGTT GGCGCTGCTG GGCTGTTACC TACGCTTGCT
GCGATCCGCG CGGGTAAAAC CATTTTGCTG GCCAATAAAG AATCACTGGT TACCTGCGGA
CGTCTGTTTA TGGACGCCGT AAAGCAGAGC AAAGCGCAAT TGTTACCGGT CGATAGCGAA
CATAACGCCA TTTTTCAGAG TTTACCGCAA CCTATCCAGC ATAATCTGGG ATACGCTGAC
CTTGAGCAAA ATGGCGTGGT GTCCATTTTA CTTACCGGGT CTGGTGGCCC TTTCCGTGAG
ACGCCATTGC GCGATTTGGC AACAATGACG CCGGATCAAG CCTGCCGTCA TCCGAACTGG
TCGATGGGGC GTAAAATTTC TGTCGATTCG GCTACCATGA TGAACAAAGG TCTGGAATAC
ATTGAAGCGC GTTGGCTGTT TAACGCCAGC GCCAGCCAGA TGGAAGTGCT GATTCACCCG
CAGTCAGTGA TTCACTCAAT GGTGCGCTAT CAGGACGGCA GTGTTCTGGC GCAGCTGGGG
GAACCGGATA TGCGTACGCC AATTGCCCAC ACCATGGCAT GGCCGAATCG CGTGAACTCT
GGCGTGAAGC CGCTCGATTT TTGCAAACTA AGTGCGTTGA CATTTGCCGC ACCGGATTAT
GATCGTTATC CATGCCTGAA ACTGGCGATG GAGGCGTTCG AACAAGGCCA GGCAGCGACG
ACAGCATTGA ATGCCGCAAA CGAAATCACC GTTGCTGCTT TTCTTGCGCA ACAAATCCGC
TTTACGGATA TCGCTGCGTT GAATTTATCC GTACTGGAAA AAATGGATAT GCGCGAACCA

CAATGTGTGG ACGATGTGTT ATCTGTTGAT GCGAACGCGC GTGAAGTCGC CAGAAAAGAG
GTGATGCGTC TCGCAAGCTG AGTCGACGAG GAGAAATTAA CCATGGCAAC CACTCATTTG
GATGTTTGCG CCGTGGTTCC GGCGGCCGGA TTTGGCCGTC GAATGCAAAC GGAATGTCCT
AAGCAATATC TCTCAATCGG TAATCAAACC ATTCTTGAAC ACTCGGTGCA TGCGCTGCTG
GCGCATCCCC GGGTGAAACG TGTCGTCATT GCCATAAGTC CTGGCGATAG CCGTTTTGCA
CAACTTCCTC TGGCGAATCA TCCGCAAATC ACCGTTGTAG ATGGCGGTGA TGAGCGTGCC
GATTCCGTGC TGGCAGGTCT GAAAGCCGCT GGCGACGCGC AGTGGGTATT GGTGCATGAC
GCCGCTCGTC CTTGTTTGCA TCAGGATGAC CTCGCGCGAT TGTTGGCGTT GAGCGAAACC
AGCCGCACGG GGGGGATCCT CGCCGCACCA GTGCGCGATA CTATGAAACG TGCCGAACCG
GGCAAAAATG CCATTGCTCA TACCGTTGAT CGCAACGGCT TATGGCACGC GCTGACGCCG
CAATTTTTCC CTCGTGAGCT GTTACATGAC TGTCTGACGC GCGCTCTAAA TGAAGGCGCG
ACTATTACCG ACGAAGCCTC GGCGCTGGAA TATTGCGGAT TCCATCCTCA GTTGGTCGAA
GGCCGTGCGG ATAACATTAA AGTCACGCGC CCGGAAGATT TGGCACTGGC CGAGTTTTAC
CTCACCCGAA CCATCCATCA GGAGAATACA TAATGCGAAT TGGACACGGT TTTGACGTAC
ATGCCTTTGG CGGTGAAGGC CCAATTATCA TTGGTGGCGT ACGCATTCCT TACGAAAAAG
GATTGCTGGC GCATTCTGAT GGCGACGTGG CGCTCCATGC GTTGACCGAT GCATTGCTTG
GCGCGGCGGC GCTGGGGGAT ATCGGCAAGC TGTTCCCGGA TACCGATCCG GCATTTAAAG
GTGCCGATAG CCGCGAGCTG CTACGCGAAG CCTGGCGTCG TATTCAGGCG AAGGGTTATA
CCCTTGGCAA CGTCGATGTC ACTATCATCG CTCAGGCACC GAAGATGTTG CCGCACATTC
CACAAATGCG CGTGTTTATT GCCGAAGATC TCGGCTGCCA TATGGATGAT GTTAACGTGA
AAGCCACTAC TACGGAAAAA CTGGGATTTA CCGGACGTGG GGAAGGGATT GCCTGTGAAG
CGGTGGCGCT ACTCATTAAG GCAACAAAAT GACTCGAGGA GGAGAAATTA ACCATGCGGA
CACAGTGGCC CTCTCCGGCA AAACTTAATC TGTTTTTATA CATTACCGGT CAGCGTGCGG
ATGGTTACCA CACGCTGCAA ACGCTGTTTC AGTTTCTTGA TTACGGCGAC ACCATCAGCA
TTGAGCTTCG TGACGATGGG GATATTCGTC TGTTAACGCC CGTTGAAGGC GTGGAACATG
AAGATAACCT GATCGTTCGC GCAGCGCGAT TGTTGATGAA AACTGCGGCA GACAGCGGGC
GTCTTCCGAC GGGAAGCGGT GCGAATATCA GCATTGACAA GCGTTTGCCG ATGGGCGGCG
GTCTCGGCGG TGGTTCATCC AATGCCGCGA CGGTCCTGGT GGCATTAAAT CATCTCTGGC
AATGCGGGCT AAGCATGGAT GAGCTGGCGG AAATGGGGCT GACGCTGGGC GCAGATGTTC
CTGTCTTTGT TCGGGGGCAT GCCGCGTTTG CCGAAGGCGT TGGTGAAATA CTAACGCCGG
TGGATCCGCC AGAGAAGTGG TATCTGGTGG CGCACCCTGG TGTAAGTATT CCGACTCCGG
TGATTTTTAA AGATCCTGAA CTCCCGCGCA ATACGCCAAA AAGGTCAATA GAAACGTTGC
TAAAATGTGA ATTCAGCAAT GATTGCGAGG TTATCGCAAG AAAACGTTTT CGCGAGGTTG
ATGCGGTGCT TTCCTGGCTG TTAGAATACG CCCCGTCGCG CCTGACTGGG ACAGGGGCCT
GTGTCTTTGC TGAATTTGAT ACAGAGTCTG AAGCCCGCCA GGTGCTAGAG CAAGCCCCGG
AATGGCTCAA TGGCTTTGTG GCGAAAGGCG CTAATCTTTC CCCATTGCAC AGAGCCATGC
TTTAAGGTAC CCAATTCGCC CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT
TTACAACGTC GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG
TTGCGCAGCC TGAATGGCGA ATGGAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT
TAAATTTTTG TTAAATCAGC TCATTTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT
ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTTGT TCCAGTTTGG AACAAGAGTC
CACTATTAAA GAACGTGGAC TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG
GCCCACTACG TGAACCATCA CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC
TAAATCGGAA CCCTAAAGGG AGCCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAACG
TGGCGAGAAA GGAAGGGAAG AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG
CGGTCACGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT
CAG

Annex I

DNA sequence of the vectorconstruct pACYClytBgcpE

ID PACYCGCLY PRELIMINARY; DNA; 5793 BP.
SQ SEQUENCE 5793 BP; 1375 A; 1506 C; 1543 G; 1369 T; 0 OTHER;

GAATTCCGGA TGAGCATTCA TCAGGCGGGC AAGAATGTGA ATAAAGGCCG GATAAAACTT
GTGCTTATTT TTCTTTACGG TCTTTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT
ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA
TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA
AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT GGTGAAAGTT
GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC
GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG TCACAGGTAT
TTATTCGGCG CAAAGTGCGT CGGGTGATGC TGCCAACTTA CTGATTTAGT GTATGATGGT
GTTTTTGAGG TGCTCCAGTG GCTTCTGTTT CTATCAGCTG TCCCTCCTGT TCAGCTACTG
ACGGGGTGGT GCGTAACGGC AAAAGCACCG CCGGACATCA GCGCTAGCGG AGTGTATACT
GGCTTACTAT GTTGGCACTG ATGAGGGTGT CAGTGAAGTG CTTCATGTGG CAGGAGAAAA
AAGGCTGCAC CGGTGCGTCA GCAGAATATG TGATACAGGA TATATTCCGC TTCCTCGCTC
ACTGACTCGC TACGCTCGGT CGTTCGACTG CGGCGAGCGG AAATGGCTTA CGAACGGGGC
GGAGATTTCC TGGAAGATGC CAGGAAGATA CTTAACAGGG AAGTGAGAGG GCCGCGGCAA
AGCCGTTTTT CCATAGGCTC CGCCCCCCTG ACAAGCATCA CGAAATCTGA CGCTCAAATC
AGTGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC GTTTCCCCCT GGCGGCTCCC
TCGTGCGCTC TCCTGTTCCT GCCTTTCGGT TTACCGGTGT CATTCCGCTG TTATGGCCGC
GTTTGTCTCA TTCCACGCCT GACACTCAGT TCCGGGTAGG CAGTTCGCTC CAAGCTGGAC
TGTATGCACG AACCCCCCGT TCAGTCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
GAGTCCAACC CGGAAAGACA TGCAAAAGCA CCACTGGCAG CAGCCACTGG TAATTGATTT
AGAGGAGTTA GTCTTGAAGT CATGCGCCGG TTAAGGCTAA ACTGAAAGGA CAAGTTTTGG
TGACTGCGCT CCTCCAAGCC AGTTACCTCG GTTCAAAGAG TTGGTAGCTC AGAGAACCTT
CGAAAAACCG CCCTGCAAGG CGGTTTTTTC GTTTTCAGAG CAAGAGATTA CGCGCAGACC
AAAACGATCT CAAGAAGATC ATCTTATTAA TCAGATAAAA TATTTCTAGA TTTCAGTGCA
ATTTATCTCT TCAAATGTAG CACCTGAAGT CAGCCCCATA CGATATAAGT TGTAATTCTC
ATGTTTGACA GCTTATCATC GATAAGCTTT AATGCGGTAG TTTATCACAG TTAAATTGCT
AACGCAGTCA GGCACCGTGT ATGAAATCTA ACAATGCGCT CATCGTCATC CTCGGCACCG
TCACCCTGGA TGCTGTAGGC ATAGGCTTGG TTATGCCGGT ACTGCCGGGC CTCTTGCGGG
ATATCGTCCA TTCCGACAGC ATCGCCAGTC ACTATGGCGT GCTGCTAGCG CTATATGCGT
TGATGCAATT TCTATGCGCA CCCGTTCTCG GAGCACTGTC CGACCGCTTT GGCCGCCGCC
CAGTCCTGCT CGCTTCGCTA CTTGGAGCCA CTATCGACTA CGCGATCATG GCGACCACAC
CCGTCCTGTG GATCCGAGGA GAAATTAACC ATGCATAACC AGGCTCCAAT TCAACGTAGA
AAATCAACAC GTATTTACGT TGGGAATGTG CCGATTGGCG ATGGTGCTCC CATCGCCGTA
CAGTCCATGA CCAATACGCG TACGACAGAC GTCGAAGCAA CGGTCAATCA AATCAAGGCG
CTGGAACGCG TTGGCGCTGA TATCGTCCGT GTATCCGTAC CGACGATGGA CGCGGCAGAA
GCGTTCAAAC TCATCAAACA GCAGGTTAAC GTGCCGCTGG TGGCTGACAT CCACTTCGAC
TATCGCATTG CGCTGAAAGT AGCGGAATAC GGCGTCGATT GTCTGCGTAT TAACCCTGGC
AATATCGGTA ATGAAGAGCG TATTCGCATG GTGGTTGACT GTGCGCGCGA TAAAAACATT
CCGATCCGTA TTGGCGTTAA CGCCGGATCG CTGGAAAAAG ATCTGCAAGA AAAGTATGGC
GAACCGACGC CGCAGGCGTT GCTGGAATCT GCCATGCGTC ATGTTGATCA TCTCGATCGC
CTGAACTTCG ATCAGTTCAA AGTCAGCGTG AAAGCGTCTG ACGTCTTCCT CGCTGTTGAG
TCTTATCGTT TGCTGGCAAA ACAGATCGAT CAGCCGTTGC ATCTGGGGAT CACCGAAGCC
GGTGGTGCGC GCAGCGGGGC AGTAAAATCC GCCATTGGTT TAGGTCTGCT GCTGTCTGAA
GGCATCGGCG ACACGCTGCG CGTATCGCTG GCGGCCGATC CGGTCGAAGA GATCAAAGTC
GGTTTCGATA TTTTGAAATC GCTGCGTATC CGTTCGCGAG GGATCAACTT CATCGCCTGC
CCGACCTGTT CGCGTCAGGA ATTTGATGTT ATCGGTACGG TTAACGCGCT GGAGCAACGC
CTGGAAGATA TCATCACTCC GATGGACGTT TCGATTATCG GCTGCGTGGT GAATGGCCCA
GGTGAGGCGC TGGTTTCTAC ACTCGGCGTC ACCGGCGGCA ACAAGAAAAG CGGCCTCTAT

GAAGATGGCG TGCGCAAAGA CCGTCTGGAC AACAACGATA TGATCGACCA GCTGGAAGCA
CGCATTCGTG CGAAAGCCAG TCAGCTGGAC GAAGCGCGTC GAATTGACGT TCAGCAGGTT
GAAAAATAAG TCGACGAGGA GAAATTAACC ATGCAGATCC TGTTGGCCAA CCCGCGTGGT
TTTTGTGCCG GGGTAGACCG CGCTATCAGC ATTGTTGAAA ACGCGCTGGC CATTTACGGC
GCACCGATAT ATGTCCGTCA CGAAGTGGTA CATAACCGCT ATGTGGTCGA TAGCTTGCGT
GAGCGTGGGG CTATCTTTAT TGAGCAGATT AGCGAAGTAC CGGACGGCGC GATCCTGATT
TTCTCCGCAC ACGGTGTTTC TCAGGCGGTA CGTAACGAAG CAAAAAGTCG CGATTTGACG
GTGTTTGATG CCACCTGTCC GCTGGTGACC AAAGTGCATA TGGAAGTCGC CCGCGCCAGT
CGCCGTGGCG AAGAATCTAT TCTCATCGGT CACGCCGGGC ACCCGGAAGT GGAAGGGACA
ATGGGCCAGT ACAGTAACCC GGAAGGGGGA ATGTATCTGG TCGAATCGCC GGACGATGTG
TGGAAACTGA CGGTCAAAAA CGAAGAGAAG CTCTCCTTTA TGACCCAGAC CACGCTGTCG
GTGGATGACA CGTCTGATGT GATCGACGCG CTGCGTAAAC GCTTCCCGAA AATTGTCGGT
CCGCGCAAAG ATGACATCTG CTACGCCACG ACTAACCGTC AGGAAGCGGT ACGCGCCCTG
GCAGAACAGG CGGAAGTTGT GTTGGTGGTC GGTTCGAAAA ACTCCTCCAA CTCCAACCGT
CTGGCGGAGC TGGCCCAGCG TATGGGCAAA CGCGCGTTTT TGATTGACGA TGCGAAAGAC
ATCCAGGAAG AGTGGGTGAA AGAGGTTAAA TGCGTCGGCG TGACTGCGGG CGCATCGGCT
CCGGATATTC TGGTGCAGAA TGTGGTGGCA CGTTTGCAGC AGCTGGGCGG TGGTGAAGCC
ATTCCGCTGG AAGGCCGTGA AGAAAACATT GTTTTCGAAG TGCCGAAAGA GCTGCGTGTC
GATATTCGTG AAGTCGATTA ACGGCCGACG CGCTGGGCTA CGTCTTGCTG GCGTTCGCGA
CGCGAGGCTG GATGGCCTTC CCCATTATGA TTCTTCTCGC TTCCGGCGGC ATCGGGATGC
CCGCGTTGCA GGCCATGCTG TCCAGGCAGG TAGATGACGA CCATCAGGGA CAGCTTCAAG
GATCGCTCGC GGCTCTTACC AGCCTAACTT CGATCACTGG ACCGCTGATC GTCACGGCGA
TTTATGCCGC CTCGGCGAGC ACATGGAACG GGTTGGCATG GATTGTAGGC GCCGCCCTAT
ACCTTGTCTG CCTCCCCGCG TTGCGTCGCG GTGCATGGAG CCGGGCCACC TCGACCTGAA
TGGAAGCCGG CGGCACCTCG CTAACGGATT CACCACTCCA AGAATTGGAG CCAATCAATT
CTTGCGGAGA ACTGTGAATG CGCAAACCAA CCCTTGGCAG AACATATCCA TCGCGTCCGC
CATCTCCAGC AGCCGCACGC GGCGCATCTC GGGCAGCGTT GGGTCCTGGC CACGGGTGCG
CATGATCGTG CTCCTGTCGT TGAGGACCCG GCTAGGCTGG CGGGGTTGCC TTACTGGTTA
GCAGAATGAA TCACCGATAC GCGAGCGAAC GTGAAGCGAC TGCTGCTGCA AAACGTCTGC
GACCTGAGCA ACAACATGAA TGGTCTTCGG TTTCCGTGTT TCGTAAAGTC TGGAAACGCG
GAAGTCCCCT ACGTGCTGCT GAAGTTGCCC GCAACAGAGA GTGGAACCAA CCGGTGATAC
CACGATACTA TGACTGAGAG TCAACGCCAT GAGCGGCCTC ATTTCTTATT CTGAGTTACA
ACAGTCCGCA CCGCTGTCCG GTAGCTCCTT CCGGTGGGCG CGGGGCATGA CTATCGTCGC
CGCACTTATG ACTGTCTTCT TTATCATGCA ACTCGTAGGA CAGGTGCCGG CAGCGCCCAA
CAGTCCCCCG GCCACGGGGC CTGCCACCAT ACCCACGCCG AAACAAGCGC CCTGCACCAT
TATGTTCCGG ATCTGCATCG CAGGATGCTG CTGGCTACCC TGTGGAACAC CTACATCTGT
ATTAACGAAG CGCTAACCGT TTTTATCAGG CTCTGGGAGG CAGAATAAAT GATCATATCG
TCAATTATTA CCTCCACGGG GAGAGCCTGA GCAAACTGGC CTCAGGCATT TGAGAAGCAC
ACGGTCACAC TGCTTCCGGT AGTCAATAAA CCGGTAAACC AGCAATAGAC ATAAGCGGCT
ATTTAACGAC CCTGCCCTGA ACCGACGACC GGGTCGAATT TGCTTTCGAA TTTCTGCCAT
TCATCCGCTT ATTATCACTT ATTCAGGCGT AGCACCAGGC GTTTAAGGGC ACCAATAACT
GCCTTAAAAA AATTACGCCC CGCCCTGCCA CTCATCGCAG TACTGTTGTA ATTCATTAAG
CATTCTGCCG ACATGGAAGC CATCACAGAC GGCATGATGA ACCTGAATCG CCAGCGGCAT
CAGCACCTTG TCGCCTTGCG TATAATATTT GCCCATGGTG AAAACGGGGG CGAAGAAGTT
GTCCATATTG GCCACGTTTA AATCAAAACT GGTGAAACTC ACCCAGGGAT TGGCTGAGAC
GAAAAACATA TTCTCAATAA ACCCTTTAGG GAAATAGGCC AGGTTTTCAC CGTAACACGC
CACATCTTGC GAATATATGT GTAGAAACTG CCGGAAATCG TCGTGGTATT CACTCCAGAG
CGATGAAAAC GTTTCAGTTT GCTCATGGAA AACGGTGTAA CAAGGGTGAA CACTATCCCA
TATCACCAGC TCACCGTCTT TCATTGCCAT ACG

//

Annex J

A and corresponding amino acid sequence of the *ispH* (formerly *lytB*) gene from *Escherichia coli*

```
         10         20         30         40         50         60         70
         |          |          |          |          |          |          |
ATGCAGATCCTGTTGGCCAACCCGCGTGGTTTTTGTGCCGGGGTAGACCGCGCTATCAGCATTGTTGAAAAC
 M  Q  I  L  L  A  N  P  R  G  F  C  A  G  V  D  R  A  I  S  I  V  E  N

        80         90        100        110        120        130        140
         |          |          |          |          |          |          |
GCGCTGGCCATTTACGGCGCACCGATATATGTCCGTCACGAAGTGGTACATAACCGCTATGTGGTCGATAGC
 A  L  A  I  Y  G  A  P  I  Y  V  R  H  E  V  V  H  N  R  Y  V  V  D  S

    150        160        170        180        190        200        210
     |          |          |          |          |          |          |
TTGCGTGAGCGTGGGGCTATCTTTATTGAGCAGATTAGCGAAGTACCGGACGGCGCGATCCTGATTTTCTCC
 L  R  E  R  G  A  I  F  I  E  Q  I  S  E  V  P  D  G  A  I  L  I  F  S

  220        230        240        250        260        270        280
   |          |          |          |          |          |          |
GCACACGGTGTTTCTCAGGCGGTACGTAACGAAGCAAAAAGTCGCGATTTGACGGTGTTTGATGCCACCTGT
 A  H  G  V  S  Q  A  V  R  N  E  A  K  S  R  D  L  T  V  F  D  A  T  C

290        300        310        320        330        340        350        360
 |          |          |          |          |          |          |          |
CCGCTGGTGACCAAAGTGCATATGGAAGTCGCCCGCGCCAGTCGCCGTGGCGAAGAATCTATTCTCATCGGT
 P  L  V  T  K  V  H  M  E  V  A  R  A  S  R  R  G  E  E  S  I  L  I  G

        370        380        390        400        410        420        430
         |          |          |          |          |          |          |
CACGCCGGGCACCCGGAAGTGGAAGGGACAATGGGCCAGTACAGTAACCCGGAAGGGGGAATGTATCTGGTC
 H  A  G  H  P  E  V  E  G  T  M  G  Q  Y  S  N  P  E  G  G  M  Y  L  V

      440        450        460        470        480        490        500
       |          |          |          |          |          |          |
GAATCGCCGGACGATGTGTGGAAACTGACGGTCAAAAACGAAGAGAAGCTCTCCTTTATGACCCAGACCACG
 E  S  P  D  D  V  W  K  L  T  V  K  N  E  E  K  L  S  F  M  T  Q  T  T

    510        520        530        540        550        560        570
     |          |          |          |          |          |          |
CTGTCGGTGGATGACACGTCTGATGTGATCGACGCGCTGCGTAAACGCTTCCCGAAAATTGTCGGTCCGCGC
 L  S  V  D  D  T  S  D  V  I  D  A  L  R  K  R  F  P  K  I  V  G  P  R

  580        590        600        610        620        630        640
   |          |          |          |          |          |          |
AAAGATGACATCTGCTACGCCACGACTAACCGTCAGGAAGCGGTACGCGCCCTGGCAGAACAGGCGGAAGTT
 K  D  D  I  C  Y  A  T  T  N  R  Q  E  A  V  R  A  L  A  E  Q  A  E  V

650        660        670        680        690        700        710        720
 |          |          |          |          |          |          |          |
GTGTTGGTGGTCGGTTCGAAAAACTCCTCCAACTCCAACCGTCTGGCGGAGCTGGCCCAGCGTATGGGCAAA
 V  L  V  V  G  S  K  N  S  S  N  S  N  R  L  A  E  L  A  Q  R  M  G  K

        730        740        750        760        770        780        790
         |          |          |          |          |          |          |
CGCGCGTTTTTGATTGACGATGCGAAAGACATCCAGGAAGAGTGGGTGAAAGAGGTTAAATGCGTCGGCGTG
 R  A  F  L  I  D  D  A  K  D  I  Q  E  E  W  V  K  E  V  K  C  V  G  V

      800        810        820        830        840        850        860
       |          |          |          |          |          |          |
ACTGCGGGCGCATCGGCTCCGGATATTCTGGTGCAGAATGTGGTGGCACGTTTGCAGCAGCTGGGCGGTGGT
 T  A  G  A  S  A  P  D  I  L  V  Q  N  V  V  A  R  L  Q  Q  L  G  G  G

    870        880        890        900        910        920        930
```

```
     |         |         |         |         |         |         |
GAAGCCATTCCGCTGGAAGGCCGTGAAGAAAACATTGTTTTCGAAGTGCCGAAAGAGCTGCGTGTCGATATT
 E  A  I  P  L  E  G  R  E  E  N  I  V  F  E  V  P  K  E  L  R  V  D  I

 940       950
   |         |
CGTGAAGTCGATTAA
 R  E  V  D  -
```

Annex K

DNA Sequence of the plasmid construct pBScyclogcpElytB2

ID PBSXICH2 PRELIMINARY; DNA; 9795 BP.
SQ SEQUENCE 9795 BP; 2351 A; 2401 C; 2770 G; 2273 T; 0 OTHER;

GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGGAGGA
GAAATTAACC ATGCATAACC AGGCTCCAAT TCAACGTAGA AAATCAACAC GTATTTACGT
TGGGAATGTG CCGATTGGCG ATGGTGCTCC CATCGCCGTA CAGTCCATGA CCAATACGCG
TACGACAGAC GTCGAAGCAA CGGTCAATCA AATCAAGGCG CTGGAACGCG TTGGCGCTGA
TATCGTCCGT GTATCCGTAC CGACGATGGA CGCGGCAGAA GCGTTCAAAC TCATCAAACA
GCAGGTTAAC GTGCCGCTGG TGGCTGACAT CCACTTCGAC TATCGCATTG CGCTGAAAGT
AGCGGAATAC GGCGTCGATT GTCTGCGTAT TAACCCTGGC AATATCGGTA ATGAAGAGCG
TATTCGCATG GTGGTTGACT GTGCGCGCGA TAAAAACATT CCGATCCGTA TTGGCGTTAA
CGCCGGATCG CTGGAAAAAG ATCTGCAAGA AAAGTATGGC GAACCGACGC CGCAGGCGTT
GCTGGAATCT GCCATGCGTC ATGTTGATCA TCTCGATCGC CTGAACTTCG ATCAGTTCAA
AGTCAGCGTG AAAGCGTCTG ACGTCTTCCT CGCTGTTGAG TCTTATCGTT TGCTGGCAAA

ACAGATCGAT CAGCCGTTGC ATCTGGGGAT CACCGAAGCC GGTGGTGCGC GCAGCGGGGC
AGTAAAATCC GCCATTGGTT TAGGTCTGCT GCTGTCTGAA GGCATCGGCG ACACGCTGCG
CGTATCGCTG GCGGCCGATC CGGTCGAAGA GATCAAAGTC GGTTTCGATA TTTTGAAATC
GCTGCGTATC CGTTCGCGAG GGATCAACTT CATCGCCTGC CCGACCTGTT CGCGTCAGGA
ATTTGATGTT ATCGGTACGG TTAACGCGCT GGAGCAACGC CTGGAAGATA TCATCACTCC
GATGGACGTT TCGATTATCG GCTGCGTGGT GAATGGCCCA GGTGAGGCGC TGGTTTCTAC
ACTCGGCGTC ACCGGCGGCA ACAAGAAAAG CGGCCTCTAT GAAGATGGCG TGCGCAAAGA
CCGTCTGGAC AACAACGATA TGATCGACCA GCTGGAAGCA CGCATTCGTG CGAAAGCCAG
TCAGCTGGAC GAAGCGCGTC GAATTGACGT TCAGCAGGTT GAAAAATAAG TCGACGAGGA
GAAATTAACC ATGCAGATCC TGTTGGCCAA CCCGCGTGGT TTTTGTGCCG GGGTAGACCG
CGCTATCAGC ATTGTTGAAA ACGCGCTGGC CATTTACGGC GCACCGATAT ATGTCCGTCA
CGAAGTGGTA CATAACCGCT ATGTGGTCGA TAGCTTGCGT GAGCGTGGGG CTATCTTTAT
TGAGCAGATT AGCGAAGTAC CGGACGGCGC GATCCTGATT TTCTCCGCAC ACGGTGTTTC
TCAGGCGGTA CGTAACGAAG CAAAAAGTCG CGATTTGACG GTGTTTGATG CCACCTGTCC
GCTGGTGACC AAAGTGCATA TGGAAGTCGC CCGCGCCAGT CGCCGTGGCG AAGAATCTAT
TCTCATCGGT CACGCCGGGC ACCCGGAAGT GGAAGGGACA ATGGGCCAGT ACAGTAACCC
GGAAGGGGGA ATGTATCTGG TCGAATCGCC GGACGATGTG TGGAAACTGA CGGTCAAAAA
CGAAGAGAAG CTCTCCTTTA TGACCCAGAC CACGCTGTCG GTGGATGACA CGTCTGATGT
GATCGACGCG CTGCGTAAAC GCTTCCCGAA AATTGTCGGT CCGCGCAAAG ATGACATCTG
CTACGCCACG ACTAACCGTC AGGAAGCGGT ACGCGCCCTG GCAGAACAGG CGGAAGTTGT
GTTGGTGGTC GGTTCGAAAA ACTCCTCCAA CTCCAACCGT CTGGCGGAGC TGGCCCAGCG
TATGGGCAAA CGCGCGTTTT TGATTGACGA TGCGAAAGAC ATCCAGGAAG AGTGGGTGAA
AGAGGTTAAA TGCGTCGGCG TGACTGCGGG CGCATCGGCT CCGGATATTC TGGTGCAGAA
TGTGGTGGCA CGTTTGCAGC AGCTGGGCGG TGGTGAAGCC ATTCCGCTGG AAGGCCGTGA
AGAAAACATT GTTTTCGAAG TGCCGAAAGA GCTGCGTGTC GATATTCGTG AAGTCGATTA
AGCGGCCGCT CTAGAACTAG TGGATCCCCC GGGCTGCAGG AATTCGAGGA GAAATTAACC
ATGTATATCG GGATAGATCT TGGCACCTCG GGCGTAAAAG TTATTTTGCT CAACGAGCAG
GGTGAGGTGG TTGCTGCGCA AACGGAAAAG CTGACCGTTT CGCGCCCGCA TCCACTCTGG
TCGGAACAAG ACCCGGAACA GTGGTGGCAG GCAACTGATC GCGCAATGAA AGCTCTGGGC
GATCAGCATT CTCTGCAGGA CGTTAAAGCA TTGGGTATTG CCGGCCAGAT GCACGGAGCA
ACCTTGCTGG ATGCTCAGCA ACGGGTGTTA CGCCCTGCCA TTTTGTGGAA CGACGGGCGC
TGTGCGCAAG AGTGCACTTT GCTGGAAGCG CGAGTTCCGC AATCGCGGGT GATTACCGGC
AACCTGATGA TGCCCGGATT TACTGCGCCT AAATTGCTAT GGGTTCAGCG GCATGAGCCG
GAGATATTCC GTCAAATCGA CAAAGTATTA TTACCGAAAG ATTACTTGCG TCTGCGTATG
ACGGGGGAGT TTGCCAGCGA TATGTCTGAC GCAGCTGGCA CCATGTGGCT GGATGTCGCA
AAGCGTGACT GGAGTGACGT CATGCTGCAG GCTTGCGACT TATCTCGTGA CCAGATGCCC
GCATTATACG AAGGCAGCGA AATTACTGGT GCTTTGTTAC CTGAAGTTGC GAAAGCGTGG
GGTATGGCGA CGGTGCCAGT TGTCGCAGGC GGTGGCGACA ATGCAGCTGG TGCAGTTGGT
GTGGGAATGG TTGATGCTAA TCAGGCAATG TTATCGCTGG GGACGTCGGG GGTCTATTTT
GCTGTCAGCG AAGGGTTCTT AAGCAAGCCA GAAAGCGCCG TACATAGCTT TTGCCATGCG
CTACCGCAAC GTTGGCATTT AATGTCTGTG ATGCTGAGTG CAGCGTCGTG TCTGGATTGG
GCCGCGAAAT TAACCGGCCT GAGCAATGTC CCAGCTTTAA TCGCTGCAGC TCAACAGGCT
GATGAAAGTG CCGAGCCAGT TTGGTTTCTG CCTTATCTTT CCGGCGAGCG TACGCCACAC
AATAATCCCC AGGCGAAGGG GGTTTTCTTT GGTTTGACTC ATCAACATGG CCCCAATGAA
CTGGCGCGAG CAGTGCTGGA AGGCGTGGGT TATGCGCTGG CAGATGGCAT GGATGTCGTG
CATGCCTGCG GTATTAAACC GCAAAGTGTT ACGTTGATTG GGGGCGGGGC GCGTAGTGAG
TACTGGCGTC AGATGCTGGC GGATATCAGC GGTCAGCAGC TCGATTACCG TACGGGGGGG
GATGTGGGGC CAGCACTGGG CGCAGCAAGG CTGGCGCAGA TCGCGGCGAA TCCAGAGAAA
TCGCTCATTG AATTGTTGCC GCAACTACCG TTAGAACAGT CGCATCTACC AGATGCGCAG
CGTTATGCCG CTTATCAGCC ACGACGAGAA ACGTTCCGTC GCCTCTATCA GCAACTTCTG
CCATTAATGG CGTAAAAGCT TGAGGAGAAA TTAACCATGA AGCAACTCAC CATTCTGGGC
TCGACCGGCT CGATTGGTTG CAGCACGCTG GACGTGGTGC GCCATAATCC CGAACACTTC
CGCGTAGTTG CGCTGGTGGC AGGCAAAAAT GTCACTCGCA TGGTAGAACA GTGCCTGGAA
TTCTCTCCCC GCTATGCCGT AATGGACGAT GAAGCGAGTG CGAAACTTCT TAAAACGATG
CTACAGCAAC AGGGTAGCCG CACCGAAGTC TTAAGTGGGC AACAAGCCGC TTGCGATATG

GCAGCGCTTG AGGATGTTGA TCAGGTGATG GCAGCCATTG TTGGCGCTGC TGGGCTGTTA
CCTACGCTTG CTGCGATCCG CGCGGGTAAA ACCATTTTGC TGGCCAATAA AGAATCACTG
GTTACCTGCG GACGTCTGTT TATGGACGCC GTAAAGCAGA GCAAAGCGCA ATTGTTACCG
GTCGATAGCG AACATAACGC CATTTTTCAG AGTTTACCGC AACCTATCCA GCATAATCTG
GGATACGCTG ACCTTGAGCA AAATGGCGTG GTGTCCATTT TACTTACCGG GTCTGGTGGC
CCTTTCCGTG AGACGCCATT GCGCGATTTG GCAACAATGA CGCCGGATCA AGCCTGCCGT
CATCCGAACT GGTCGATGGG GCGTAAAATT TCTGTCGATT CGGCTACCAT GATGAACAAA
GGTCTGGAAT ACATTGAAGC GCGTTGGCTG TTTAACGCCA GCGCCAGCCA GATGGAAGTG
CTGATTCACC CGCAGTCAGT GATTCACTCA ATGGTGCGCT ATCAGGACGG CAGTGTTCTG
GCGCAGCTGG GGGAACCGGA TATGCGTACG CCAATTGCCC ACACCATGGC ATGGCCGAAT
CGCGTGAACT CTGGCGTGAA GCCGCTCGAT TTTTGCAAAC TAAGTGCGTT GACATTTGCC
GCACCGGATT ATGATCGTTA TCCATGCCTG AAACTGGCGA TGGAGGCGTT CGAACAAGGC
CAGGCAGCGA CGACAGCATT GAATGCCGCA AACGAAATCA CCGTTGCTGC TTTTCTTGCG
CAACAAATCC GCTTTACGGA TATCGCTGCG TTGAATTTAT CCGTACTGGA AAAAATGGAT
ATGCGCGAAC CACAATGTGT GGACGATGTG TTATCTGTTG ATGCGAACGC GCGTGAAGTC
GCCAGAAAAG AGGTGATGCG TCTCGCAAGC TGAGTCGACG AGGAGAAATT AACCATGGCA
ACCACTCATT TGGATGTTTG CGCCGTGGTT CCGGCGGCCG GATTTGGCCG TCGAATGCAA
ACGGAATGTC CTAAGCAATA TCTCTCAATC GGTAATCAAA CCATTCTTGA ACACTCGGTG
CATGCGCTGC TGGCGCATCC CCGGGTGAAA CGTGTCGTCA TTGCCATAAG TCCTGGCGAT
AGCCGTTTTG CACAACTTCC TCTGGCGAAT CATCCGCAAA TCACCGTTGT AGATGGCGGT
GATGAGCGTG CCGATTCCGT GCTGGCAGGT CTGAAAGCCG CTGGCGACGC GCAGTGGGTA
TTGGTGCATG ACGCCGCTCG TCCTTGTTTG CATCAGGATG ACCTCGCGCG ATTGTTGGCG
TTGAGCGAAA CCAGCCGCAC GGGGGGGATC CTCGCCGCAC CAGTGCGCGA TACTATGAAA
CGTGCCGAAC CGGGCAAAAA TGCCATTGCT CATACCGTTG ATCGCAACGG CTTATGGCAC
GCGCTGACGC CGCAATTTTT CCCTCGTGAG CTGTTACATG ACTGTCTGAC GCGCGCTCTA
AATGAAGGCG CGACTATTAC CGACGAAGCC TCGGCGCTGG AATATTGCGG ATTCCATCCT
CAGTTGGTCG AAGGCCGTGC GGATAACATT AAAGTCACGC GCCCGGAAGA TTTGGCACTG
GCCGAGTTTT ACCTCACCCG AACCATCCAT CAGGAGAATA CATAATGCGA ATTGGACACG
GTTTTGACGT ACATGCCTTT GGCGGTGAAG GCCCAATTAT CATTGGTGGC GTACGCATTC
CTTACGAAAA AGGATTGCTG GCGCATTCTG ATGGCGACGT GGCGCTCCAT GCGTTGACCG
ATGCATTGCT TGGCGCGGCG GCGCTGGGGG ATATCGGCAA GCTGTTCCCG GATACCGATC
CGGCATTTAA AGGTGCCGAT AGCCGCGAGC TGCTACGCGA AGCCTGGCGT CGTATTCAGG
CGAAGGGTTA TACCCTTGGC AACGTCGATG TCACTATCAT CGCTCAGGCA CCGAAGATGT
TGCCGCACAT TCCACAAATG CGCGTGTTTA TTGCCGAAGA TCTCGGCTGC CATATGGATG
ATGTTAACGT GAAAGCCACT ACTACGGAAA AACTGGGATT TACCGGACGT GGGGAAGGGA
TTGCCTGTGA AGCGGTGGCG CTACTCATTA AGGCAACAAA ATGACTCGAG GAGGAGAAAT
TAACCATGCG GACACAGTGG CCCTCTCCGG CAAAACTTAA TCTGTTTTTA TACATTACCG
GTCAGCGTGC GGATGGTTAC CACACGCTGC AAACGCTGTT TCAGTTTCTT GATTACGGCG
ACACCATCAG CATTGAGCTT CGTGACGATG GGGATATTCG TCTGTTAACG CCCGTTGAAG
GCGTGGAACA TGAAGATAAC CTGATCGTTC GCGCAGCGCG ATTGTTGATG AAAACTGCGG
CAGACAGCGG GCGTCTTCCG ACGGGAAGCG GTGCGAATAT CAGCATTGAC AAGCGTTTGC
CGATGGGCGG CGGTCTCGGC GGTGGTTCAT CCAATGCCGC GACGGTCCTG GTGGCATTAA
ATCATCTCTG GCAATGCGGG CTAAGCATGG ATGAGCTGGC GGAAATGGGG CTGACGCTGG
GCGCAGATGT TCCTGTCTTT GTTCGGGGGC ATGCCGCGTT TGCCGAAGGC GTTGGTGAAA
TACTAACGCC GGTGGATCCG CCAGAGAAGT GGTATCTGGT GGCGCACCCT GGTGTAAGTA
TTCCGACTCC GGTGATTTTT AAAGATCCTG AACTCCCGCG CAATACGCCA AAAAGGTCAA
TAGAAACGTT GCTAAAATGT GAATTCAGCA ATGATTGCGA GGTTATCGCA AGAAAACGTT
TTCGCGAGGT TGATGCGGTG CTTTCCTGGC TGTTAGAATA CGCCCCGTCG CGCCTGACTG
GGACAGGGGC CTGTGTCTTT GCTGAATTTG ATACAGAGTC TGAAGCCCGC CAGGTGCTAG
AGCAAGCCCC GGAATGGCTC AATGGCTTTG TGGCGAAAGG CGCTAATCTT TCCCCATTGC
ACAGAGCCAT GCTTTAAGGT ACCCAATTCG CCCTATAGTG AGTCGTATTA CGCGCGCTCA
CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC
CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC
CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGAAAT TGTAAGCGTT AATATTTTGT
TAAAATTCGC GTTAAATTTT TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG

GCAAAATCCC TTATAAATCA AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT
GGAACAAGAG TCCACTATTA AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT
ATCAGGGCGA TGGCCCACTA CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT
GCCGTAAAGC ACTAAATCGG AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA
AGCCGGCGAA CGTGGCGAGA AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC
TGGCAAGTGT AGCGGTCACG CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC
TACAGGGCGC GTCAG

//

Annex L

DNA and corresponding amino acid sequence of the *ispG* gene (fragment) from *Arabidopsis thaliana*

```
          10         20         30         40         50         60         70
          |          |          |          |          |          |          |
>AAGACGGTGAGAAGGAAGACTCGTACTGTTATGGTTGGAAATGTCGCCCTTGGAAGCGAACATCCGATAAGG
  K  T  V  R  R  K  T  R  T  V  M  V  G  N  V  A  L  G  S  E  H  P  I  R

        80         90        100        110        120        130        140
        |          |          |          |          |          |          |
ATTCAAACGATGACTACTTCGGATACAAAAGATATTACTGGAACTGTTGATGAGGTTATGAGAATAGCGGAT
  I  Q  T  M  T  T  S  D  T  K  D  I  T  G  T  V  D  E  V  M  R  I  A  D

     150        160        170        180        190        200        210
      |          |          |          |          |          |          |
AAAGGAGCTGATATTGTAAGGATAACTGTTCAAGGGAAGAAAGAGGCGGATGCGTGCTTTGAAATAAAAGAT
  K  G  A  D  I  V  R  I  T  V  Q  G  K  K  E  A  D  A  C  F  E  I  K  D

   220        230        240        250        260        270        280
    |          |          |          |          |          |          |
AAACTCGTTCAGCTTAATTACAATACACCGCTGGTTGCAGGTATTCATTTTGCCCCTACTGTAGCCTTACGA
  K  L  V  Q  L  N  Y  N  T  P  L  V  A  G  I  H  F  A  P  T  V  A  L  R

290        300        310        320        330        340        350        360
 |          |          |          |          |          |          |          |
GTCGCTGAATGCTTTGACAAGATCCGTGTCAACCCCGGAAATTTTGCGGACAGGCGGGCCCAGTTTGAGACG
  V  A  E  C  F  D  K  I  R  V  N  P  G  N  F  A  D  R  R  A  Q  F  E  T

        370        380        390        400        410        420        430
        |          |          |          |          |          |          |
ATAGATTATACAGAAGATGAATATCAGAAAGAACTCCAGCATATCGAGCAGGTCTTCACTCCTTTGGTTGAG
  I  D  Y  T  E  D  E  Y  Q  K  E  L  Q  H  I  E  Q  V  F  T  P  L  V  E

     440        450        460        470        480        490        500
      |          |          |          |          |          |          |
AAATGCAAAAAGTACGGGAGAGCAATGCGTATTGGGACAAATCATGGAAGTCTTTCTGACCGTATCATGAGC
  K  C  K  K  Y  G  R  A  M  R  I  G  T  N  H  G  S  L  S  D  R  I  M  S

   510        520        530        540        550        560        570
    |          |          |          |          |          |          |
TATTACGGGGATTCTCCCCGAGGAATGGTTGAATCTGCGTTTGAGTTTGCAAGAATATGTCGGAAATTAGAC
  Y  Y  G  D  S  P  R  G  M  V  E  S  A  F  E  F  A  R  I  C  R  K  L  D

  580        590        600        610        620        630        640
   |          |          |          |          |          |          |
TATCACAACTTTGTTTTCTCAATGAAAGCGAGCAACCCAGTGATCATGGTCCAGGCGTACCGTTTACTTGTG

  Y  H  N  F  V  F  S  M  K  A  S  N  P  V  I  M  V  Q  A  Y  R  L  L  V

650        660        670        680        690        700        710        720
 |          |          |          |          |          |          |          |
GCTGAGATGTATGTTCATGGATGGGATTATCCTTTGCATTTGGGAGTTACTGAGGCAGGAGAAGGCGAAGAT
  A  E  M  Y  V  H  G  W  D  Y  P  L  H  L  G  V  T  E  A  G  E  G  E  D

        730        740        750        760        770        780        790
        |          |          |          |          |          |          |
GGACGGATGAAATCTGCGATTGGAATTGGGACGCTTCTTCAGGACGGGCTCGGTGACACAACAAGAGTTTCA
  G  R  M  K  S  A  I  G  I  G  T  L  L  Q  D  G  L  G  D  T  T  R  V  S

     800        810        820        830        840        850        860
      |          |          |          |          |          |          |
CTGACGGAGCCACCAGAAGAGGAGATAGATCCCTGCAGGCGATTGGCTAACCTCGGGACAAAAGCTGCCAAA
```

```
  L  T  E  P  P  E  E  E  I  D  P  C  R  R  L  A  N  L  G  T  K  A  A  K

    870        880        890        900        910        920        930
     |          |          |          |          |          |          |
CTTCAACAAGGCGCTGCACCGTTTGAAGAAAAGCATAGGCATTACTTTGATTTTCAGCGTCGGACGGGTGAT
  L  Q  Q  G  A  A  P  F  E  E  K  H  R  H  Y  F  D  F  Q  R  R  T  G  D

   940        950        960        970        980        990       1000
    |          |          |          |          |          |          |
CTACCTGTACAAAAAGAGGGAGAAGAGGTTGATTACAGAAATGTCCTTCACCGTGATGGTTCTGTTCTGATG
  L  P  V  Q  K  E  G  E  E  V  D  Y  R  N  V  L  H  R  D  G  S  V  L  M

1010       1020       1030       1040       1050       1060       1070       1080
  |          |          |          |          |          |          |          |
TCGATTTCTCTGGATCAACTAAAGGCACCTGAACTCCTCTACAGATCACTCGCCACAAAGCTTGTCGTGGGT
  S  I  S  L  D  Q  L  K  A  P  E  L  L  Y  R  S  L  A  T  K  L  V  V  G

        1090       1100       1110       1120       1130       1140       1150
          |          |          |          |          |          |          |
ATGCCATTCAAGGATCTGGCAACTGTTGATTCAATCTTATTAAGAGAGCTACCGCCTGTAGATGATCAAGTG
  M  P  F  K  D  L  A  T  V  D  S  I  L  L  R  E  L  P  P  V  D  D  Q  V

     1160       1170       1180       1190       1200       1210       1220
       |          |          |          |          |          |          |
GCTCGTTTGGCTCTCAAACGGTTGATTGATGTCAGTATGGGAGTTATAGCACCTTTATCAGAGCAACTAACA
  A  R  L  A  L  K  R  L  I  D  V  S  M  G  V  I  A  P  L  S  E  Q  L  T

   1230       1240       1250       1260       1270       1280       1290
     |          |          |          |          |          |          |
AAGCCATTGCCCAATGCCATGGTTCTTGTCAACCTCAAGGAACTATCTGGTGGCGCTTACAAGCTTCTCCCT
  K  P  L  P  N  A  M  V  L  V  N  L  K  E  L  S  G  G  A  Y  K  L  L  P

  1300       1310       1320       1330       1340       1350       1360
    |          |          |          |          |          |          |
GAAGGTACACGCTTGGTTGTCTCTCTACGAGGCGATGAGCCTTACGAGGAGCTTGAAATACTCAAAAACATT
  E  G  T  R  L  V  V  S  L  R  G  D  E  P  Y  E  E  L  E  I  L  K  N  I

1370       1380       1390       1400       1410       1420       1430       1440
  |          |          |          |          |          |          |          |
GATGCTACTATGATTCTCCATGATGTACCTTTCACTGAAGACAAAGTTAGCAGAGTACATGCAGCTCGGAGG
  D  A  T  M  I  L  H  D  V  P  F  T  E  D  K  V  S  R  V  H  A  A  R  R

        1450       1460       1470       1480       1490       1500       1510
          |          |          |          |          |          |          |
CTATTCGAGTTCTTATCCGAGAATTCAGTTAACTTTCCTGTTATTCATCGCATAAACTTCCCAACCGGAATC
  L  F  E  F  L  S  E  N  S  V  N  F  P  V  I  H  R  I  N  F  P  T  G  I

     1520       1530       1540       1550       1560       1570       1580
       |          |          |          |          |          |          |
CACAGAGACGAATTGGTGATTCATGCAGGGACATATGCTGGAGGCCTTCTTGTGGATGGACTAGGTGATGGC
  H  R  D  E  L  V  I  H  A  G  T  Y  A  G  G  L  L  V  D  G  L  G  D  G

   1590       1600       1610       1620       1630       1640       1650
     |          |          |          |          |          |          |
GTAATGCTCGAAGCACCTGACCAAGATTTTGATTTTCTTAGGAATACTTCCTTCAACTTATTACAAGGATGC
  V  M  L  E  A  P  D  Q  D  F  D  F  L  R  N  T  S  F  N  L  L  Q  G  C

  1660       1670       1680       1690       1700       1710       1720
    |          |          |          |          |          |          |
AGAATGCGTAACACTAAGACGGAATATGTATCGTGCCCGTCTTGTGGAAGAACGCTTTTCGACTTGCAAGAA
  R  M  R  N  T  K  T  E  Y  V  S  C  P  S  C  G  R  T  L  F  D  L  Q  E

1730       1740       1750       1760       1770       1780       1790       1800
  |          |          |          |          |          |          |          |
ATCAGCGCCGAGATCCGAGAAAAGACTTCCCATTTACCTGGCGTTTCGATCGCAATCATGGGATGCATTGTG
  I  S  A  E  I  R  E  K  T  S  H  L  P  G  V  S  I  A  I  M  G  C  I  V

        1810       1820       1830       1840       1850       1860       1870
```

```
         |         |         |         |         |         |         |
AATGGACCAGGAGAAATGGCAGATGCTGATTTCGGATATGTAGGTGGTTCTCCCGGAAAAATCGACCTTTAT
 N  G  P  G  E  M  A  D  A  D  F  G  Y  V  G  G  S  P  G  K  I  D  L  Y

    1880      1890      1900      1910      1920      1930      1940
      |         |         |         |         |         |         |
GTCGGAAAGACGGTGGTGAAGCGTGGGATAGCTATGACGGAGGCAACAGATGCTCTGATCGGTCTGATCAAA
 V  G  K  T  V  V  K  R  G  I  A  M  T  E  A  T  D  A  L  I  G  L  I  K

  1950      1960      1970      1980
    |         |         |         |
GAACATGGTCGTTGGGTCGACCCGCCCGTGGCTGATGAGTAG
 E  H  G  R  W  V  D  P  P  V  A  D  E  -
```

Annex M

DNA and corresponding amino acid sequence of the *ispG* (forrmly *gcpE*) gene of *Arabidopsis thaliana*

```
ATGGCGACTGGAGTATTGCCAGCTCCGGTTTCTGGGATCAAGATACCGGATTCGAAAGTC   60
 M  A  T  G  V  L  P  A  P  V  S  G  I  K  I  P  D  S  K  V    20

GGGTTTGGTAAAAGCATGAATCTTGTGAGAATTTGTGATGTTAGGAGTCTAAGATCTGCT  120
 G  F  G  K  S  M  N  L  V  R  I  C  D  V  R  S  L  R  S  A    40

AGGAGAAGAGTTTCGGTTATCCGGAATTCAAACCAAGGCTCTGATTTAGCTGAGCTTCAA  180
 R  R  R  V  S  V  I  R  N  S  N  Q  G  S  D  L  A  E  L  Q    60

CCTGCATCCGAAGGAAGCCCTCTCTTAGTGCCAAGACAGAAATATTGTGAATCATTGCAT  240
 P  A  S  E  G  S  P  L  L  V  P  R  Q  K  Y  C  E  S  L  H    80

AAGACGGTGAGAAGGAAGACTCGTACTGTTATGGTTGGAAATGTCGCCCTTGGAAGCGAA  300
 K  T  V  R  R  K  T  R  T  V  M  V  G  N  V  A  L  G  S  E   100

CATCCGATAAGGATTCAAACGATGACTACTTCGGATACAAAAGATATTACTGGAACTGTT  360
 H  P  I  R  I  Q  T  M  T  T  S  D  T  K  D  I  T  G  T  V   120

GATGAGGTTATGAGAATAGCGGATAAAGGAGCTGATATTGTAAGGATAACTGTTCAAGGG  420
 D  E  V  M  R  I  A  D  K  G  A  D  I  V  R  I  T  V  Q  G   140

AAGAAAGAGGCGGATGCGTGCTTTGAAATAAAAGATAAACTCGTTCAGCTTAATTACAAT  480
 K  K  E  A  D  A  C  F  E  I  K  D  K  L  V  Q  L  N  Y  N   160

ACACCGCTGGTTGCAGGTATTCATTTTGCCCCTACTGTAGCCTTACGAGTCGCTGAATGC  540
 T  P  L  V  A  G  I  H  F  A  P  T  V  A  L  R  V  A  E  C   180

TTTGACAAGATCCGTGTCAACCCCGGAAATTTTGCGGACAGGCGGGCCCAGTTTGAGACG  600
 F  D  K  I  R  V  N  P  G  N  F  A  D  R  R  A  Q  F  E  T   200

ATAGATTATACAGAAGATGAATATCAGAAAGAACTCCAGCATATCGAGCAGGTCTTCACT  660
 I  D  Y  T  E  D  E  Y  Q  K  E  L  Q  H  I  E  Q  V  F  T   220

CCTTTGGTTGAGAAATGCAAAAAGTACGGGAGAGCAATGCGTATTGGGACAAATCATGGA  720
 P  L  V  E  K  C  K  K  Y  G  R  A  M  R  I  G  T  N  H  G   240

AGTCTTTCTGACCGTATCATGAGCTATTACGGGGATTCTCCCCGAGGAATGGTTGAATCT  780
 S  L  S  D  R  I  M  S  Y  Y  G  D  S  P  R  G  M  V  E  S   260

GCGTTTGAGTTTGCAAGAATATGTCGGAAATTAGACTATCACAACTTTGTTTTCTCAATG  840
 A  F  E  F  A  R  I  C  R  K  L  D  Y  H  N  F  V  F  S  M   280

AAAGCGAGCAACCCAGTGATCATGGTCCAGGCGTACCGTTTACTTGTGGCTGAGATGTAT  900
 K  A  S  N  P  V  I  M  V  Q  A  Y  R  L  L  V  A  E  M  Y   300

GTTCATGGATGGGATTATCCTTTGCATTTGGGAGTTACTGAGGCAGGAGAAGGCGAAGAT  960
 V  H  G  W  D  Y  P  L  H  L  G  V  T  E  A  G  E  G  E  D   320

GGACGGATGAAATCTGCGATTGGAATTGGGACGCTTCTTCAGGACGGGCTCGGTGACACA 1020
```

```
 G  R  M  K  S  A  I  G  I  G  T  L  L  Q  D  G  L  G  D  T

ACAAGAGTTTCACTGACGGAGCCACCAGAAGAGGAGATAGATCCCTGCAGGCGATTGGCT 1080
 T  R  V  S  L  T  E  P  P  E  E  E  I  D  P  C  R  R  L  A   360

AACCTCGGGACAAAAGCTGCCAAACTTCAACAAGGCGCTGCACCGTTTGAAGAAAAGCAT 1140
 N  L  G  T  K  A  A  K  L  Q  Q  G  A  A  P  F  E  E  K  H   380


AGGCATTACTTTGATTTTCAGCGTCGGACGGGTGATCTACCTGTACAAAAAGAGGGAGAA 1200
 R  H  Y  F  D  F  Q  R  R  T  G  D  L  P  V  Q  K  E  G  E   400

GAGGTTGATTACAGAAATGTCCTTCACCGTGATGGTTCTGTTCTGATGTCGATTTCTCTG 1260
 E  V  D  Y  R  N  V  L  H  R  D  G  S  V  L  M  S  I  S  L   420

GATCAACTAAAGGCACCTGAACTCCTCTACAGATCACTCGCCACAAAGCTTGTCGTGGGT 1320
 D  Q  L  K  A  P  E  L  L  Y  R  S  L  A  T  K  L  V  V  G   440

ATGCCATTCAAGGATCTGGCAACTGTTGATTCAATCTTATTAAGAGAGCTACCGCCTGTA 1380
 M  P  F  K  D  L  A  T  V  D  S  I  L  L  R  E  L  P  P  V   460

GATGATCAAGTGGCTCGTTTGGCTCTCAAACGGTTGATTGATGTCAGTATGGGAGTTATA 1440
 D  D  Q  V  A  R  L  A  L  K  R  L  I  D  V  S  M  G  V  I   480

GCACCTTTATCAGAGCAACTAACAAAGCCATTGCCCAATGCCATGGTTCTTGTCAACCTC 1500
 A  P  L  S  E  Q  L  T  K  P  L  P  N  A  M  V  L  V  N  L   500

AAGGAACTATCTGGTGGCGCTTACAAGCTTCTCCCTGAAGGTACACGCTTGGTTGTCTCT 1560
 K  E  L  S  G  G  A  Y  K  L  L  P  E  G  T  R  L  V  V  S   520

CTACGAGGCGATGAGCCTTACGAGGAGCTTGAAATACTCAAAAACATTGATGCTACTATG 1620
 L  R  G  D  E  P  Y  E  E  L  E  I  L  K  N  I  D  A  T  M   540

ATTCTCCATGATGTACCTTTCACTGAAGACAAAGTTAGCAGAGTACATGCAGCTCGGAGG 1680
 I  L  H  D  V  P  F  T  E  D  K  V  S  R  V  H  A  A  R  R   560

CTATTCGAGTTCTTATCCGAGAATTCAGTTAACTTTCCTGTTATTCATCGCATAAACTTC 1740
 L  F  E  F  L  S  E  N  S  V  N  F  P  V  I  H  R  I  N  F   580

CCAACCGGAATCCACAGAGACGAATTGGTGATTCATGCAGGGACATATGCTGGAGGCCTT 1800
 P  T  G  I  H  R  D  E  L  V  I  H  A  G  T  Y  A  G  G  L   600

CTTGTGGATGGACTAGGTGATGGCGTAATGCTCGAAGCACCTGACCAAGATTTTGATTTT 1860
 L  V  D  G  L  G  D  G  V  M  L  E  A  P  D  Q  D  F  D  F   620

CTTAGGAATACTTCCTTCAACTTATTACAAGGATGCAGAATGCGTAACACTAAGACGGAA 1920
 L  R  N  T  S  F  N  L  L  Q  G  C  R  M  R  N  T  K  T  E   640

TATGTATCGTGCCCGTCTTGTGGAAGAACGCTTTTCGACTTGCAAGAAATCAGCGCCGAG 1980
 Y  V  S  C  P  S  C  G  R  T  L  F  D  L  Q  E  I  S  A  E   660

ATCCGAGAAAAGACTTCCCATTTACCTGGCGTTTCGATCGCAATCATGGGATGCATTGTG 2040
 I  R  E  K  T  S  H  L  P  G  V  S  I  A  I  M  G  C  I  V   680

AATGGACCAGGAGAAATGGCAGATGCTGATTTCGGATATGTAGGTGGTTCTCCCGGAAAA 2100
```

```
 N  G  P  G  E  M  A  D  A  D  F  G  Y  V  G  G  S  P  G  K   700

ATCGACCTTTATGTCGGAAAGACGGTGGTGAAGCGTGGGATAGCTATGACGGAGGCAACA 2160
 I  D  L  Y  V  G  K  T  V  V  K  R  G  I  A  M  T  E  A  T   720

GATGCTCTGATCGGTCTGATCAAAGAACATGGTCGTTGGGTCGACCCGCCCGTGGCTGAT 2220
 D  A  L  I  G  L  I  K  E  H  G  R  W  V  D  P  P  V  A  D   740

GAGTAG 2226
 E  -   741
```

Annex N

cDNA sequence of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IspH) from *Arabidopsis thaliana*

```
ATGGCTGTTGCGCTCCAATTCAGCCGATTATGCGTTCGACCGGATACTTTCGTGCGGGAGAATCATCTCTCT   72
 M  A  V  A  L  Q  F  S  R  L  C  V  R  P  D  T  F  V  R  E  N  H  L  S    24

GGATCCGGATCTCTCCGCCGCCGGAAAGCTTTATCAGTCCGGTGCTCGTCTGGCGATGAGAACGCTCCTTCG  144
 G  S  G  S  L  R  R  R  K  A  L  S  V  R  C  S  S  G  D  E  N  A  P  S    48

CCATCGGTGGTGATGGACTCCGATTTCGACGCCAAGGTGTTCCGTAAGAACTTGACGAGAAGCGATAATTAC  216
 P  S  V  V  M  D  S  D  F  D  A  K  V  F  R  K  N  L  T  R  S  D  N  Y    72

AATCGTAAAGGGTTCGGTCATAAGGAGGAGACACTCAAGCTCATGAATCGAGAGTACACCAGTGATATATTG  288
 N  R  K  G  F  G  H  K  E  E  T  L  K  L  M  N  R  E  Y  T  S  D  I  L    96

GAGACACTGAAAACAAATGGGTATACTTATTCTTGGGGAGATGTTACTGTGAAACTCGCTAAAGCATATGGT  360
 E  T  L  K  T  N  G  Y  T  Y  S  W  G  D  V  T  V  K  L  A  K  A  Y  G   120

TTTTGCTGGGGTGTTGAGCGTGCTGTTCAGATTGCATATGAAGCACGAAAGCAGTTTCCAGAGGAGAGGCTT  432
 F  C  W  G  V  E  R  A  V  Q  I  A  Y  E  A  R  K  Q  F  P  E  E  R  L   144

TGGATTACTAACGAAATCATTCATAACCCGACCGTCAATAAGAGGTTGGAAGATATGGATGTTAAAATTATT  504
 W  I  T  N  E  I  I  H  N  P  T  V  N  K  R  L  E  D  M  D  V  K  I  I   168

CCGGTTGAGGATTCAAAGAAACAGTTTGATGTAGTAGAGAAAGATGATGTGGTTATCCTTCCTGCGTTTGGA  576
 P  V  E  D  S  K  K  Q  F  D  V  V  E  K  D  D  V  V  I  L  P  A  F  G   192

GCTGGTGTTGACGAGATGTATGTTCTTAATGATAAAAAGGTGCAAATTGTTGACACGACTTGTCCTTGGGTG  648
 A  G  V  D  E  M  Y  V  L  N  D  K  K  V  Q  I  V  D  T  T  C  P  W  V   216

ACAAAGGTCTGGAACACGGTTGAGAAGCACAAGAAGGGGGAATACACATCAGTAATCCATGGTAAATATAAT  720
 T  K  V  W  N  T  V  E  K  H  K  K  G  E  Y  T  S  V  I  H  G  K  Y  N   240

CATGAAGAGACGATTGCAACTGCGTCTTTTGCAGGAAAGTACATCATTGTAAAGAACATGAAAGAGGCAAAT  792
 H  E  E  T  I  A  T  A  S  F  A  G  K  Y  I  I  V  K  N  M  K  E  A  N   264

TACGTTTGTGATTACATTCTCGGTGGCCAATACGATGGATCTAGCTCCACAAAAGAGGAGTTCATGGAGAAA  864
 Y  V  C  D  Y  I  L  G  G  Q  Y  D  G  S  S  S  T  K  E  E  F  M  E  K   288

TTCAAATACGCAATTTCGAAGGGTTTCGATCCCGACAATGACCTTGTCAAAGTTGGTATTGCAAACCAAACA  936
 F  K  Y  A  I  S  K  G  F  D  P  D  N  D  L  V  K  V  G  I  A  N  Q  T   312

ACGATGCTAAAGGGAGAAACAGAGGAGATAGGAAGATTACTCGAGACAACAATGATGCGCAAGTATGGAGTG 1008
 T  M  L  K  G  E  T  E  E  I  G  R  L  L  E  T  T  M  M  R  K  Y  G  V   336

GAAAATGTAAGCGGACATTTCATCAGCTTCAACACAATATGCGACGCTACTCAAGAGCGACAAGACGCAATC 1080
 E  N  V  S  G  H  F  I  S  F  N  T  I  C  D  A  T  Q  E  R  Q  D  A  I   360

TATGAGCTAGTGGAAGAGAAGATTGACCTCATGCTAGTGGTTGGCGGATGGAATTCAAGTAACACCTCTCAC 1152
 Y  E  L  V  E  E  K  I  D  L  M  L  V  V  G  G  W  N  S  S  N  T  S  H   384

CTTCAGGAAATCTCAGAGGCACGGGGAATCCCATCTTACTGGATCGATAGTGAGAAACGGATAGGACCTGGG 1224
 L  Q  E  I  S  E  A  R  G  I  P  S  Y  W  I  D  S  E  K  R  I  G  P  G   408

AATAAAATAGCCTATAAGCTCCACTATGGAGAACTGGTCGAGAAGGAAAACTTTCTCCCAAAGGGACCAATA 1296
 N  K  I  A  Y  K  L  H  Y  G  E  L  V  E  K  E  N  F  L  P  K  G  P  I   432

ACAATCGGTGTGACATCAGGTGCATCAACCCCGGATAAGGTCGTGGAAGATGCTTTGGTGAAGGTGTTCGAC 1368
 T  I  G  V  T  S  G  A  S  T  P  D  K  V  V  E  D  A  L  V  K  V  F  D   456

ATTAAACGTGAAGAGTTATTGCAGCTGGCTTGA                                         1398
 I  K  R  E  E  L  L  Q  L  A  -                                           466
```